SANKHYĀ

THE INDIAN JOURNAL OF STATISTICS

SANKHYĀ, Series A, Vol. 24, Part 1, 1962

CONTENTS

(All rights reserved)

INDIAN STATISTICAL INSTITUTE

The Indian Statistical Institute is a Society registered in April 1932 under the Indian Societies Registration Act (Act No. XXI of 1860). Since 1 April 1960, it has been declared as an Institution of National Importance by the Indian Statistical Institute Act No. 57 of 1959.

Extracts from the Memorandum of Association

II. *The objects of the Institute shall be:—*

(1) *To promote the study and dissemination of knowledge of and research on Statistics and other subjects relating to planning for national development and social welfare.*

(2) *To provide for and undertake the collection of information, investigations, projects, and operational research for purposes of planning and the improvement of efficiency of management and production.*

(3) *To undertake any other activities which in the opinion of the Council may be usefully carried out by the Institute in the furtherance of national development and social welfare.*

IV. *The membership of the Institute and its constituent bodies shall be open to all persons irrespective of sex, nationality, race, creed or class.*

V. *The Institute shall not distribute any part of its income or profit as dividend, gift, or bonus or in any other way among its members, provided nevertheless that office-bearers, workers and other employees shall not be debarred from receiving remuneration or participating in grants because of their also being members of the Institute.*

The Institute consists of Ordinary and Life Members, Institutional Members, Fellows, Honorary Members and Patrons. The supreme control including the power of making rules is vested in the members in General Meeting assembled. The Chairman and other office-bearers are elected once every two years. The management of the affairs of the Institute is in the hands of a Council elected annually by the members of the Institute.

Facilities are provided for student and sessional membership.

Membership fee, Rs. 25/- (foreign Rs. 30/-) per year, Institutional Membership fee, Rs. 100/- per year include subscription to one series of *Sankhyā: The Indian Journal of Statistics.*

Detailed information and Membership forms may be obtained on application to the Indian Statistical Institute, 203 Barrackpore Trunk Road, Calcutta-35.

PAPERS

2

BIOMETRIKA

Volume 48, Parts 3 and 4   **December, 1961**

CONTENTS

THE ANNALS OF MATHEMATICAL STATISTICS

(Founded by H. C. Carver)

The Official Journal of the Institute of Mathematical Statistics

CONTENTS OF VOLUME 33, NO. 1, MARCH, 1962

ECONOMETRICA

Vol. 30, No. 1　　　　　　　　　　　　　　　　　　**January, 1962**

CONTENTS

SANKHYĀ

THE INDIAN JOURNAL OF STATISTICS

Edited by : P. C. MAHALANOBIS

THE SIMULTANEOUS DISTRIBUTION OF CORRELATION COEFFICIENTS

By SIR RONALD A. FISHER

Division of Mathematical Statistics, C.S.I.R.O.

SUMMARY. From any sample in t variables an inter-related system of $\frac{1}{2}t(t-1)$ correlations r_{ij} can be calculated. Their simultaneous distribution depends only on the corresponding system of true correlations in the hypothetical multinormal population sampled, and in the following paper is expressed explicitly in terms of them.

1. If a sample of N be available from the simultaneous distribution of t correlated variates, direct multiplication of the N independent elements of frequency gives the expression

$$(2\pi)^{-\frac{1}{2}tN}(\sigma_1 \sigma_2 \ldots \sigma_t)^{-N}|\rho_{ij}|^{-\frac{1}{2}N}$$

$$\exp\left[-\tfrac{1}{2}S\left\{\frac{(x_i-\mu_i)^2}{\sigma_i^2}\rho^{ii}+\ldots+\frac{2(x_i-\mu_i)(x_j-\mu_j)}{\sigma_i\sigma_j}\rho^{ij}+\ldots\right\}\right]$$

$$dx_{11}\ldots dx_{1N}dx_{21}\ldots dx_{2N}dx_{t1}\ldots dx_{tN} \qquad \ldots \quad (1)$$

where S stands for summation over the N sets of observations; $\sigma_1, \sigma_2, \ldots, \sigma_t$ for the standard deviations in the population sampled, $|\rho_{ij}|$ for the determinant of the population correlations, and ρ^{ij} for the elements of the reciprocal of the matrix of population correlations ρ_{ij}.

From this primitive product, the simultaneous sampling distributions are to be inferred of (a) the sample means,

$$x = \frac{1}{N}\,S(x), \qquad \ldots \quad (2)$$

1

of each variate, (b) of the estimates of the variances,

$$s^2 = \frac{1}{N-1} S(x-\bar{x})^2 \qquad \qquad \ldots \quad (3)$$

and finally, of the estimated coefficients of correlation

$$r_{ij} = \frac{S(x_i-\bar{x}_i)(x_j-\bar{x}_j)}{s_i s_j (N-1)} \qquad \qquad \ldots \quad (4)$$

and from this the marginal simultaneous distribution of the r_{ij}.

This program was carried out in 1915 for the case of two variates, and the distribution of the estimate r in terms of the true value ρ only, has been available in practical use for many years. For more than two variables the problem seems to have been left in a quite incomplete form.

As in the bivariate case it is sufficiently clear that the estimates s_i and r_{ij} are distributed independently of the empirical means, $\bar{x}_i$, in a distribution which does not involve the true means, μ_i, so that on integrating out all means, the marginal distribution is the same as that conditional on specific values. Equally, the $t(t-1)/2$ statistics r_{ij} will be distributed independently not only of the means, but also of the standard deviations.

The factor representing the simultaneous distribution of the $\bar{x}_i$ is easily found and removed by considering the transformation (rotation) of each variate effected by the orthogonal matrix G,

where
$$X_i' = GX_i$$

and X_i stands for the vector $(x_{i1}, x_{i2}, \ldots, x_{iN})$

for if
$$X_{i1}' = S_j(x_{ij})/\sqrt{N} = \bar{x}_i \sqrt{N}$$

the t variables
$$X_{i1}',$$

will be distributed in exactly the same simultaneous distribution as the original t variables, so that the omission of this one factor leaves a product equivalent to that appropriate to a simple sample of $(N-1)$ i.e.,

$$(2\pi)^{-\frac{1}{2}t(N-1)}(\sigma_1\sigma_2 \ldots \sigma_t)^{-(N-1)}|\rho_{ij}|^{-\frac{1}{2}(N-1)}$$

$$\exp\left[-\frac{N-1}{2}\left\{\frac{s_i^2}{\sigma_i^2}\rho^{ii}+\ldots+\frac{2r_{ij}s_is_j}{\sigma_i\sigma_j}\rho^{ij}+\ldots\right\}\right]dv \qquad \ldots \quad (5)$$

in which s_i^2 has been substituted for

$$\left.\begin{array}{l} \dfrac{1}{N-1}\{S(x_i^2)-N\bar{x}_i^2\} \\[2em] \end{array}\right\}$$

and $\qquad r_{ij}s_is_j$ for $\dfrac{1}{N-1}\{S(x_ix_j)-N\bar{x}_i\bar{x}_j\}$ $\qquad \ldots \quad (6)$

and finally, dv for the element of volume in $t(N-1)$ dimensions.

2.   The second important step was taken by Wishart (1928) by expressing the element of volume in terms of the quantities

$$\xi_{ii} = S(x_i - \bar{x}_i)^2$$
$$\xi_{ij} = S(x_i - \bar{x}_i)(x_j - \bar{x}_j).$$

Regarding these quadratic forms as coordinates, Wishart showed that

$$dv = \frac{\pi^{t(2N-t-1)/4}}{\left(\frac{N-3}{2}\right)! \dots \left(\frac{N-t-2}{2}\right)!} \, V^{(N-t-2)/2} \, d\xi_{11} \dots d\xi_{tt} \qquad \dots \quad (7)$$

where V stands for the determinant of the ξ_{ij}.

To obtain the distribution explicitly in terms of the s_i and the r_{ij}, by substituting

$$\left. \begin{aligned} \xi_{ii} &= (N-1)s_i^2 \\ \xi_{ij} &= (N-1)r_{ij}s_i s_j, \end{aligned} \right\} \qquad \dots \quad (8)$$

we may note

$$V = s_1^2 \dots s_t^2 (N-1)^t |r_{ij}|$$

where $|r_{ij}|$ stands for determinant of the r_{ij}, and

$$\frac{\partial(\xi_{11} \dots \xi_{tt})}{\partial(r_{ij}, \, s_i)} = 2^t (s_1 \dots s_t)^t (N-1)^{t(t+1)/2},$$

giving

$$\frac{(N-1)^{\frac{1}{2}t(N-1)} \pi^{-t(t-1)/4} |\rho_{ij}|^{-\frac{1}{2}(N-1)}}{\left(\frac{N-3}{2}\right)! \dots \left(\frac{N-t-2}{2}\right)! \, 2^{\frac{1}{2}t(N-3)}} \left(\frac{s_1 \dots s_t}{\sigma_1 \dots \sigma_t}\right)^{N-2}$$

$$\exp\left[-\frac{N-1}{2} \left\{ \frac{s_i^2}{\sigma_i^2} \rho^{ii} + \dots + \frac{2r_{ij}s_i s_j}{\sigma_i \sigma_j} \rho^{ij} + \dots \right\} \right]$$

$$|r_{ij}|^{\frac{1}{2}(N-t-2)} \frac{ds_1}{\sigma_1} \dots \frac{ds_t}{\sigma_t} \Pi(dr_{ij}). \qquad \dots \quad (9)$$

3.   So far we have the direct result of substituting the s_i and the r_{ij} in Wishart's distribution when the numerical factors are restored, which is not quite easy with the form as left by Wishart.   Further simplification flows from the substitution

$$u_i = \frac{s_i}{\sigma_i} \sqrt{(N-1)\rho^{ii}} \qquad \dots \quad (10)$$

which leads to

$$\frac{\pi^{-t(t-1)/4} \, 2^{-t(N-3)/2}}{\left(\frac{N-3}{2}\right)! \dots \left(\frac{N-t-2}{2}\right)!} \frac{|\rho_{ij}|^{-\frac{1}{2}(N-1)}}{(\rho^{11}\rho^{22} \dots \rho^{tt})^{\frac{1}{2}(N-1)}} |r_{ij}|^{(N-t-2)/2} \Pi(dr_{ij})$$

$$\exp\left\{ -\frac{1}{2} (u_1^2 + \dots + u_t^2 - 2\gamma_{12}u_1 u_2 - \dots) \right\}$$

$$(u_1 \dots u_t)^{N-2} du_1 \dots du_t \qquad \dots \quad (11)$$

where
$$\gamma_{ij} = -\frac{\rho^{ij} r_{ij}}{\sqrt{\rho^{ii} \rho^{jj}}}.$$
... (12)

It may be observed that the range of s_i is from 0 to ∞, that u_i is s_i multiplied by a positive constant, and that

$$\int_0^\infty \cdots \int_0^\infty \exp\{-\tfrac{1}{2}(u_1^2 + \ldots + u_t^2 - 2\gamma_{12} u_1 u_2 - \ldots\}(u_1 \ldots u_t)^{N-2} du_1 \ldots du_t \quad \ldots \ (13)$$

depends only on the coefficients γ_{ij}, and on N.

Writing
$$F_{N-2}(\gamma_{ij})$$

for this integral, the simultaneous distribution of the values r_{ij}, in number $t(t-1)/2$, is

$$\frac{\pi^{-t(t-1)/4} \, 2^{-t(N-3)/2}}{\left(\frac{N-3}{2}\right)! \ldots \left(\frac{N-t-2}{2}\right)!} \frac{|\rho_{ij}|^{-\frac{1}{2}(N-1)}}{(\rho^{11} \rho^{22} \ldots \rho^{tt})^{\frac{1}{2}(N-1)}}$$

$$|r_{ij}|^{(N-t-2)/2} F_{n-2}(\gamma_{ij}) \, \Pi(dr_{ij}) \qquad \ldots \ (14)$$

when the variation of $s_1, s_2, \ldots, s_t$ is eliminated.

Denoting
$$\rho_{ij}^* = \frac{\rho^{ij}}{\sqrt{\rho^{ii} \rho^{jj}}}$$

and therefore
$$\gamma_{ij} = -r_{ij} \rho_{ij}^*$$

the expression (14) may be written

$$\frac{\pi^{-t(t-1)/4} \, 2^{-t(N-3)/2}}{\left(\frac{N-3}{2}\right)! \ldots \left(\frac{N-t-2}{2}\right)!} |\rho_{ij}^*|^{(N-1)/2} |r_{ij}|^{(N-t-2)/2} F_{N-2}(\gamma_{ij}) \, \Pi(dr_{ij}) \ldots \ (15)$$

Geometrically, if ρ_{ij} is the cosine of the angular distance between two points i and j on a hypersphere, then ρ_{ij}^* is the cosine with reversed sign of the dihedral angle opposite to these, or to the cosine of the angular distance between corresponding points of the polar figure.

Wilks (1944, p. 120) proposed that this elimination should be carried out by expanding the exponential as in (9) in powers of its exponent, a quadratic in s_i, and by eliminating these variates by integration. This, however, is not a feasible path for more than two variates. The case of two variates had been solved nearly 30 years earlier by quite a different approach. Wilks makes no specific proposal for more than two variables.

4. *Analytic notes.* (a) *Certain definite integrals.*

From equation (15) can be derived a number of cognate identities. If, for any values of N and t all the true correlations ρ_{ij} are zero, it follows that all γ_{ij} are zero also. Consequently the determinant

$$|\rho_{ij}^{2}|$$

is replaced by unity, and the function F by

$$2^{t(N-3)/2} \left(\frac{N-3}{2} \ ! \right)^{t}$$

then putting
$$N = t+2$$
we find the generalised volume obtained by integration with respect to r_{ij} over all possible values i.e., over the closed region within which the determinant remains positive, in the expression

$$\frac{0 ! \, \frac{1}{2} ! \, \dots \, \frac{t-1}{2} !}{\left(\frac{t-1}{2} ! \right)^{t}} \ \pi^{t(t-1)/4} . \qquad \dots \ (16)$$

For any value of t the number of dimensions of the generalised volume is $\frac{1}{2}t(t-1)$; the following table gives algebraic and numerical value for moderate values of t.

TABLE. GENERALISED VOLUME OF REGIONS OF
INTEGRATION FOR CORRELATION
COEFFICIENTS

t	dimensions	generalised volume	numerical value
2	1	2	2.
3	3	$\pi^2/2$	4.9348
4	6	$32\pi^2/27$	11.6973
5	10	$3\pi^6/128$	22.5326
6	15	$2^{13}\pi^6/3^4 5^5$	31.114
7	21	$5\pi^{12}/3^4 2^{11}$	27.858
8	28	$2^{24}\pi^{12}/3^4 5^6 7^7$	14.877
9	36	$5^2 7\pi^{20}/3^4 2^{32}$	4.4115
10	45	$2^{46}\pi^{20}/3^{32} 5^7 7^8$	.68227

More generally if $N = t+2+k$ and $|r_{ij}| = D^2$,

$$\int \dots \int D^k dr_{12} \dots dr_{t-1,t} = \frac{\frac{k}{2} ! \, \frac{k+1}{2} ! \dots \frac{k+t-1}{2} !}{\left(\frac{k+t-1}{2} \ ! \right)^{t}} \ \pi^{t(t-1)/4} \qquad \dots \ (17)$$

giving the integrals of all powers of D, and the means of all powers over this region.

(b) *Derivatives of $F_{N-2}(\gamma_{ij} = 0)$.*

The function $F_{N-2}(\gamma_{ij})$ when γ_{ij} are all zero may be regarded as a special case of the more general function

$$\int_0^\infty u_1^{N-2+s_1} \, e^{-u_1^2/2} \, du_1 \int_0^\infty u_2^{N-2+s_2} \, e^{-u_2^2/2} \, du_2 \dots$$

for all t variables u; this product is easily evaluated as the product of

$$2^{(N+s_1-3)/2} \; \frac{N+s_1-3}{2}! \;\; 2^{(N+s_2-3)/2} \; \frac{N+s_2-3}{2}! \; \dots \qquad \qquad \dots \text{ (18)}$$

Consequently, if a_{ij} are any positive integers, or zero, and if

$$\Sigma_j \, a_{ij} = A_i$$

so that

$$\Sigma_i \, A_i = 2 \, \Sigma \Sigma_{i>j} \, a_{ij},$$

then the differential coefficient

$$\Pi \, \frac{\partial^{a_{ij}}}{\partial \gamma_{ij}^{a_{ij}}} \, F_{N-2}(\gamma_{ij} = 0)$$

will be

$$2^{(N+A_1-2)/2} \; \frac{N+A_1-2}{2}! \;\; 2^{(N+A_2-2)/2} \; \frac{N+A_2-2}{2}! \dots \qquad \dots \text{ (19)}$$

so giving completely the differential coefficients of $F_{N-2}(\gamma_{ij} = 0)$ with respect to the $\frac{1}{2}t(t-1)$ variables γ_{ij}.

In particular the first differential coefficients are all equal, namely,

$$2^{N-2} \left(\frac{N-2}{2}! \right)^2 2^{(t-2)(N-3)/2} \left(\frac{N-3}{2}! \right)^{t-2}$$

or

$$2. \, \frac{\left(\dfrac{N-2}{2}! \right)^2}{\left(\dfrac{N-3}{2}! \right)^2} \, F$$

so that

$$\frac{\partial}{\partial (\gamma_{ij})} \log F = 2 \, \frac{\left(\dfrac{N-2}{2}! \right)^2}{\left(\dfrac{N-3}{2}! \right)^2} \text{ for all } i,j. \qquad \dots \text{ (20)}$$

Similarly

$$\frac{\partial^2}{\partial \gamma_{ij}^2} \log F = (N-1)^2 - 4 \, \frac{\left(\dfrac{N-2}{2}! \right)^4}{\left(\dfrac{N-3}{2}! \right)^4}$$

$$\frac{\partial^2}{\partial \gamma_{ij} \partial \gamma_{ik}} \log F = 2(N-1) \, \frac{\left(\dfrac{N-2}{2}! \right)^2}{\left(\dfrac{N-3}{2}! \right)^2} - 4 \, \frac{\left(\dfrac{N-2}{2}! \right)^4}{\left(\dfrac{N-3}{2}! \right)^4} \qquad \dots \text{ (21)}$$

and

$$\frac{\partial^2}{\partial \gamma_{ij} \partial \gamma_{kl}} \log F = 0$$

where separate terms refer to different partitions of the aggregate a_{ij}. Log F is then expressible as a multiple power series in γ_{ij}. In connection with these ratios of factorials, it may be useful to note that

$$2\left(\frac{N-2}{2}!\right)^2 \div \left(\frac{N-3}{2}!\right)^2.$$

may be expanded in the form

$$N-\frac{3}{2}+\frac{1}{8(N-3/2)}-\frac{9}{128(N-3/2)^3}+\frac{153}{1024(N-3/2)^5}-\cdots \qquad \cdots \quad (22)$$

so that

$$\frac{\partial}{\partial \gamma_{ij}}\log F \sim N-\frac{3}{2}$$

$$\frac{\partial^2}{\partial \gamma_{ij}^2}\log F \sim N-\frac{3}{2}$$

$$\frac{\partial^2}{\partial \gamma_{ij}\partial \gamma_{ik}}\log F \sim \frac{1}{2}\left(N-\frac{7}{4}\right)$$

all being nearly equal, apart from a simple coefficient.

(c) *Geometrical interpretation.*

The expression $F_0(\gamma_{ij})$ has a rather simple geometrical interpretation. When, for example, $t=3$ we may consider the transformation of the original coordinates u, v, w into new coordinates u', v', w', according to homogeneous linear equations

$$u' = \lambda_1 u + \mu_1 v + \nu_1 w$$
$$v' = \lambda_2 u + \mu_2 v + \nu_2 w$$
$$w' = \lambda_3 u + \mu_3 v + \nu_3 w.$$

Then
$$u'^2+v'^2+w'^2 = u^2+v^2+w^2-2\gamma_{12}uv-2\gamma_{13}uw-2\gamma_{23}vw$$

if

$$\lambda_1^2+\lambda_2^2+\lambda_3^2 = 1, \qquad \mu_1\nu_1+\mu_2\nu_2+\mu_3\nu_3 = -\gamma_{23}$$
$$\mu_1^2+\mu_2^2+\mu_3^2 = 1, \qquad \lambda_1\nu_1+\lambda_2\nu_2+\lambda_3\nu_3 = -\gamma_{13}$$
$$\nu_1^2+\nu_2^2+\nu_3^2 = 1, \qquad \lambda_1\mu_1+\lambda_2\mu_2+\lambda_3\mu_3 = -\gamma_{12},$$

and the bounding edge $v=0$, $w=0$ has direction cosines

$$\lambda_1, \lambda_2, \lambda_3,$$

and so with the others, so that the cosines of the angles between pairs of such edges are

$$\mu_1\nu_1+\mu_2\nu_2+\mu_3\nu_3 = -\gamma_{23}, \quad \text{etc.}$$

Noting that

$$\int_0^\infty e^{-r^2/2}r^2 dr = \sqrt{2}\,(\tfrac{1}{2})! = \sqrt{\pi/2}$$

or for t dimensions
$$\int_0^\infty e^{-r^2/2}r^{t-1}dr = 2^{(t-2)/2}\,\frac{t-2}{2}! \qquad \cdots \quad (23)$$

it is seen that $F_0(\gamma_{ij})$ is for three dimensions, the area of the spherical triangle the sides of which have cosines $(-\gamma_{ij})$, multiplied by $\sqrt{\pi/2}$, and for t dimensions the generalised volume of the corresponding hyperspherical figure, multiplied by

$$2^{(t-2)/2}\cdot\frac{t-2}{2}!$$

and in each case divided by

$$\partial(u', v', w')/\partial(u, v, w),$$

or by the determinant

$$\begin{vmatrix} \lambda_1 & \mu_1 & \nu_1 \\ \lambda_2 & \mu_2 & \nu_2 \\ \lambda_3 & \mu_3 & \nu_3 \end{vmatrix}$$

of which the square is

$$\begin{vmatrix} 1 & -\gamma_{12} & -\gamma_{13} \\ -\gamma_{12} & 1 & -\gamma_{23} \\ -\gamma_{13} & -\gamma_{23} & 1 \end{vmatrix}$$

If therefore this is written D^2, then in general

$$F_0(\gamma_{ij}) = 2^{(t-2)/2} \frac{t-2}{2}! \frac{V}{D} \qquad \qquad \dots \ (24)$$

where V is the generalised volume of the figure defined by $-\gamma_{ij}$, and D is the 'genera-lised sine' of the solid angle it subtends at the centre of the hypersphere, or the volume defined by the t unit vectors.

When t is even $F_1 (\gamma_{ij})$ can be derived from F_0 directly by differentiation with respect to chosen variables, e.g., for $t = 4$, as

$$\frac{\partial^2}{\partial \gamma_{12} \partial \gamma_{34}} F_0(\gamma_{ij}) \qquad \qquad \dots \ (25)$$

If t is odd, the suffix can be increased in this way only by steps of two; so when $t=3$

$$F_2(\gamma_{ij}) = \frac{\partial^3}{\partial \gamma_{12} \partial \gamma_{23} \partial \gamma_{31}} F_0(\gamma_{ij}) \qquad \qquad \dots \ (26)$$

A general form for F_1, when t is odd, is thus to be desired. In the case $t = 3$, fairly direct integration may be used to show that, if $\cos \theta_1 = \gamma_{23}$, etc., $0 \leqslant \theta \leqslant \pi$, then

$$F_1 D^2 = 1 + \frac{1}{D^2} \Sigma \frac{\theta_1}{\sin \theta_1} \{\gamma_{23}(1 - \gamma_{23}^2) + \gamma_{31}(\gamma_{12} + \gamma_{23}\gamma_{31}) + \gamma_{12}(\gamma_{31} + \gamma_{12}\gamma_{23})\}. \quad \dots \ (27)$$

REFERENCES

FISHER, R. A. (1915) : The frequency distribution of the correlation coefficient in samples from an indefi-
nitely large population. *Biom.*, **10**, 507-521.

WISHART, J. (1928) : The generalised product-moment distribution in samples from a normal multivariate
population. *Biom.*, **20a**, 32-52.

WILKS, S. S. (1944) : *Mathematical Statistics*. Princeton University Press.

RESOLVABLE INCOMPLETE BLOCK DESIGNS WITH TWO REPLICATIONS*

By R. C. BOSE

and

K. R. NAIR**

Institute of Statistics, University of North Carolina

SUMMARY. A class of resolvable incomplete block designs with two replications is given here. This includes the known two replicate designs, which are either the sample square or rectangular lattices and their extensions. The designs in this class are not necessarily PBIB designs but their duals are PBIB designs with three associate classes. Nevertheless the treatments in these designs have an association scheme with a maximum of seven associate classes.

A list of all the designs with $k \leqslant 10$ belonging to this class has been prepared and methods of construction of designs explained. The necessary formulae for analysis with recovery of interblock information are derived.

1. INTRODUCTION

Incomplete block designs become most useful when the number of treatments or varieties is very large. In such cases it is seldom that the research worker will have experimental material for more than a small number of replications. It is therefore of practical importance to make available to him a repertoire of designs involving as few replications as possible.

For a systematic search for such designs, their construction and analysis, it is advisable to go step by step from designs requiring two replications to those requiring three replications, four replications, and so on.

In this paper, we shall consider two-replicate designs only. There are two main groups of Incomplete Block designs already investigated, namely, the *Balanced* Incomplete Block designs due to Yates (1936) and the *Partially Balanced* Incomplete Block (PBIB) designs developed by the authors (1939) and whose definition was slightly modified by Nair and Rao (1942).

In virtue of the fact that the number of replications of each treatment in a Balanced Incomplete Block design cannot be less than the number of plots per block and since the latter cannot be less than two, the search for two-replicate Balanced Incomplete Block designs has to be confined to the case where there are two plots per block. There is only one set of such balanced designs, namely, the one where every pair of treatments is made to constitute one block. In the familiar notation :

$$\lambda(v-1) = r(k-1)$$

the only two-replicate balanced design we can construct is for the trivial case,

$$v = 3, \ k = 2, \ r = 2, \ b = 3, \ \lambda = 1.$$

Since *Partially Balanced* Incomplete Block designs are free from the restriction $r \geqslant k$ a number of PBIB designs can be constructed for which $r = 2$. Bose (1951) made an exhaustive study of two-replicate PBIB designs having two associate classes. Nair (1950, 1951) gave some examples of two-replicate PBIB designs

---

involving three or four associate classes. Among the designs considered by them there is a sub-class where the blocks can be arranged in two separate replication groups. This is a feature which, if taken advantage of while laying out the experiment, can help in recovering inter-block information, and in thereby improving the efficiency of the design. Such designs will be called *resolvable* designs.

The only resolvable two-replicate PBIB designs found so far are basically the simple square and rectangular lattices involving p^2 and $p(p-1)$ treatments respectively in blocks of p and $(p-1)$ plots. The former involves two and the latter four associate classes of treatment comparisons. By replacing each treatment by a set of q treatments in each of these lattices we can evolve two-replicate PBIB designs with three and five associate classes for p^2q and $p(p-1)q$ treatments in blocks of pq and $(p-1)q$ treatments respectively.

In this paper we shall evolve a new class of resolvable Incomplete Block designs having only two replications. In general, these designs have seven associate classes of treatment comparisons, but in particular cases they may be less than seven. This new class of designs does *not in general* belong to the category of PBIB designs defined by Bose, Nair and Rao, but the dual designs of this class, i.e., the designs obtained by interchanging treatments and blocks, are PBIB designs with three associate classes. The four PBIB designs mentioned in the last paragraph are special cases of this new class of designs.

The method of analysis of the new class of designs with recovery of interblock information is simple and straightforward. The necessary formulae have been derived.

A list of all the designs with $k \leqslant 10$ has been prepared and their methods of construction explained.

2. THE DESIGN

(1) Consider the $u \times u$ incidence matrix of a *Symmetrical* Balanced Incomplete Block design where u treatments are replicated r times so that there are u blocks of r plots each and

$$\lambda(u-1) = r(r-1). \qquad \ldots \ (2.1)$$

Let the matrix be denoted by

$$N = (n_{ij}) \qquad \ldots \ (2.2)$$

where n_{ij}, the element at the junction of i-th row and j-th column, is either 0 or 1.

The following conditions are satisfied by the elements of matrix (2.2)

$$\sum_{i=1}^{u} n_{ij} = \sum_{i=1}^{u} n_{ij}^2 = r \qquad (j = 1, 2, ..., u) \qquad \ldots \ (2.3)$$

$$\sum_{j=1}^{u} n_{ij} = \sum_{j=1}^{u} n_{ij}^2 = r \qquad (i = 1, 2, ..., u) \qquad \ldots \ (2.4)$$

$$\sum_{j=1}^{u} n_{hj} n_{ij} = \lambda \qquad (h \neq i = 1, 2, ..., u) \qquad \ldots \ (2.5)$$

$$\sum_{i=1}^{u} n_{ij} n_{ih} = \lambda \qquad (j \neq h = 1, 2, ..., u). \qquad \ldots \ (2.6)$$

(2)   In the matrix (2.2) we shall replace every n_{ij} having value 1 by a set of p treatments and every n_{ij} having value 0 by a set of q treatments.   No treatment is repeated inside the matrix.   The total number of treatments in any row or column of the matrix will be

$$k = rp + (u-r)q \qquad \qquad \dots \quad (2.7)$$

and in the whole matrix :

$$v = ku. \qquad \qquad \dots \quad (2.8)$$

To construct the design we take all the k treatments in the same row of the matrix to form a single block and thus build up u blocks from the u rows to form one replication.   We do likewise with the k treatments in the same column of the matrix and form u blocks of a second replication from the u columns.   The design will therefore have two resolvable replications with $2u$ blocks of k plots each.

(3)   We will consider a few examples at this stage.

Example 1.   Let $u = 3$, $r = 2$.   Then $\lambda = 1$ and

$$N = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}$$

Let $p = 2$, $q = 1$.   Representing the l-th treatment in the cell (ij) by t_{ij}^l, the scheme of the design is :

t_{11}^1 t_{11}^2	t_{12}^1 t_{12}^2	t_{13}^1
t_{21}^1	t_{22}^1 t_{22}^2	t_{23}^1 t_{23}^2
t_{31}^1 t_{31}^2	t_{32}^1	t_{33}^1 t_{33}^2

Re-identifying the treatments and calling them 1, 2, ..., 15, we could write the scheme in a simpler way as

1,2	3,4	5
6	7,8	9,10
11,12	13	14,15

The design is then :

	replication I						replication II				
block number	treatment number						block number	treatment number			
(1)	1	2	3	4	5		(4)	1	2	6	11	12
(2)	6	7	8	9	10		(5)	3	4	7	8	13
(3)	11	12	13	14	15		(6)	5	9	10	14	15

We thus have a resolvable two-replicate design for 15 treatments in six blocks of 5 plots each.

Example 2. Let $u = r$. In other words we are dealing with symmetrical complete blocks, so that the n_{ij} in the incidence matrix is 1 in every cell and $\lambda = r$. Putting $p = 1$, we get the simple square lattice for $v = u^2$, $k = u$. When $p > 1$, we get the extended square lattice design : $v = u^2 p$, $k = up$.

Example 3. Let $u = r+1$. Then $\lambda = r-1$. By putting $p = 1$, $q = 0$, we get the simple rectangular lattice for $v = u(u-1)$, $k = (u-1)$. If $p > 1$, $q = 0$, we get the extended rectangular lattice design : $v = u(u-1)p$, $k = (u-1)p$.

3. Estimation of treatment effects

(1) Let us denote the m_{ij} treatments of cell (ij) by $ij(1)$, $ij(2)$, ..., $ij(m_{ij})$ and their effects by

$$t_{ij}^{(1)}, t_{ij}^{(2)}, ..., t_{ij}^{(m_{ij})},$$

where $m_{ij} = p$ or q according as $n_{ij} = 1$ or 0.

Let
$$R_i(t) = \sum_{j=1}^{u} \sum_{l=1}^{m_{ij}} t_{ij}^{(l)} \qquad \qquad ...\ (3.1)$$

$$= \text{Sum of the treatment effects for row } i$$

$$C_j(t) = \sum_{i=1}^{u} \sum_{l=1}^{m_{ij}} t_{ij}^{(l)} \qquad \qquad ...\ (3.2)$$

$$= \text{Sum of the treatment effects for column } j.$$

For the sake of unique estimation of the effects, we shall make the usual assumption that $\Sigma\ \Sigma\ \Sigma\ t_{ij}^{(l)} = 0$, or, that

$$\sum_{i=1}^{u} R_i(t) = \sum_{j=1}^{u} C_j(t) = 0. \qquad \qquad ...\ (3.3)$$

For treatment $ij(l)$, let $x_{ij}^{(l)}$ be the observed value of the character under study in the first replication and $y_{ij}^{(l)}$ that in the second replication. We shall call these replications x- and y-replications.

Let
$$R_i(x) = \sum_{j=1}^{u} \sum_{l=1}^{m_{ij}} x_{ij}^{(l)} \qquad \qquad ...\ (3.4)$$

$$C_j(x) = \sum_{i=1}^{u} \sum_{l=1}^{m_{ij}} x_{ij}^{(l)}. \qquad \qquad ...\ (3.5)$$

Let
$$T(x) = \sum_{i=1}^{u} R_i(x) = \sum_{j=1}^{u} C_j(x)$$

$$R_i(y) = \sum_{j=1}^{u} \sum_{l=1}^{m_{ij}} y_{ij}^{(l)} \qquad \qquad ...\ (3.6)$$

$$C_j(y) = \sum_{i=1}^{u} \sum_{l=1}^{m_{ij}} y_{ij}^{(l)}. \qquad \qquad ...\ (3.7)$$

Let $$T(y) = \Sigma R_i(y) = \Sigma C_j(y).$$

It will be noted that $R_i(x)$ is the total value of x for the i-th block of the x-replication and that $C_j(y)$ is the total value of y for the j-th block of the y-replication.

Let $Q_{ij}^{(l)}$ be the total observed value of the character under study for the two replications of treatment $ij(l)$ minus the sum of the corresponding block means. Hence,

$$Q_{ij}^{(l)} = x_{ij}^{(l)} + y_{ij}^{(l)} - \frac{1}{k}\{R_i(x) + C_j(y)\}. \qquad \dots \text{(3.8)}$$

Let us introduce a quantity $Q_{ij}^{\prime(l)}$ defined by

$$Q_{ij}^{\prime(l)} = \frac{1}{k}\{R_i(x) + C_j(y)\} - \frac{1}{v}\{T(x) + T(y)\}. \qquad \dots \text{(3.9)}$$

Let $$R_i(Q) = \overset{u}{\underset{j=1}{\Sigma}} \overset{m_{ij}}{\underset{l=1}{\Sigma}} Q_{ij}^{(l)} \; ; \; R_i(Q') = \overset{u}{\underset{j=1}{\Sigma}} \overset{m_{ij}}{\underset{l=1}{\Sigma}} Q_{ij}^{\prime(l)}. \qquad \dots \text{(3.10)}$$

$$C_j(Q) = \overset{u}{\underset{i=1}{\Sigma}} \overset{m_{ij}}{\underset{l=1}{\Sigma}} Q_{ij}^{(l)} \; ; \; C_j(Q') = \overset{u}{\underset{i=1}{\Sigma}} \overset{m_{ij}}{\underset{l=1}{\Sigma}} Q_{ij}^{\prime(l)}. \qquad \dots \text{(3.11)}$$

(2) The normal equation corresponding to the treatment $ij(l)$ may be written as

$$k\{wQ_{ij}^{(l)} + w'Q_{ij}^{\prime(l)}\} = 2kwt_{ij}^{(l)} - (w-w')\{R_i(t) + C_j(t)\} \qquad \dots \text{(3.12)}$$

where w and w' are reciprocals of the intra- and inter-block variance per plot.

Summing up over j and l

$$k\{wR_i(Q) + w'R_i(Q')\} = k(w+w')R_i(t) - (w-w')\overset{u}{\underset{j=1}{\Sigma}} m_{ij}C_j(t). \qquad \dots \text{(3.13)}$$

Summing up over i and l

$$k\{wC_j(Q) + w'C_j(Q')\} = k(w+w')C_j(t) - (w-w')\overset{u}{\underset{i=1}{\Sigma}} m_{ij}R_i(t). \qquad \dots \text{(3.14)}$$

Substituting for $C_j(t)$ from (3.14) in (3.13) and putting

$$\frac{w-w'}{w+w'} = \gamma, \qquad \dots \text{(3.15)}$$

we get, $C_{i1}R_1(t) + C_{i2}R_2(t) + \dots + C_{iu}R_u(t) = k\{wR_i(Q) + w'R_i(Q')\}$ $\dots$ (3.16)

$$+ \gamma \overset{u}{\underset{j=1}{\Sigma}} m_{ij}\{wC_j(Q) + w'C_j(Q')\}$$

where $$C_{ii} = k(w+w')\left\{1 - \left(\frac{\gamma}{k}\right)^2 (m_{i1}^2 + m_{i2}^2 + \dots + m_{iu}^2)\right\} \qquad \dots \text{(3.17)}$$

$$C_{ih} = -\frac{\gamma^2}{k}(w+w')\{m_{i1}m_{h1} + m_{i2}m_{h2} + \dots + m_{iu}m_{hu}\} \qquad \dots \text{(3.18)}$$

$$i \neq h.$$

Scince $m_{ij} = p$ or q according as $n_{ij} = 1$ or 0, (3.17) and (3.18) reduce to

$$C_{ii} = k(w+w')\left[1-\left(\frac{\gamma}{k}\right)^2 rp^2+(u-r)q^2\right] \qquad \dots \text{(3.19)}$$

$$C_{ih} = -\frac{\gamma^2}{k}(w+w')\{\lambda p^2+2(r-\lambda)pq+(u-2r+\lambda)q^2\}. \qquad \dots \text{(3.20)}$$

Using the condition (3.3) we can simplify (3.16) to the form

$$(C_{ii}-C_{ih})R_i(t) = k\{wR_i(Q)+w'R_i(Q')\}+\gamma\sum_{j=1}^{u} m_{ij}\{wC_j(Q)+w'C_j(Q')\}. \qquad \dots \text{(3.21)}$$

Similarly, we get

$$(C_{ii}-C_{ih})C_j(t) = k\{wC_j(Q)+w'C_j(Q')\}+\gamma\sum_{i=1}^{u} m_{ij}\{wR_i(Q)+w'R_i(Q')\}. \qquad \dots \text{(3.22)}$$

The value of $(C_{ii}-C_{ih})$ occurring in (3.21) and (3.22) can be simplified to the form

$$C_{ii}-C_{ih} = \frac{1}{k}(w+w')\{k^2-(r-\lambda)(p-q)^2\gamma^2\}. \qquad \dots \text{(3.23)}$$

(3) Substituting (3.21) and (3.22) in (3.12), we have

$$t_{ij}^{*(l)} = \frac{1}{2w}(wQ_{ij}^{(l)}+w'Q_{ij}'^{(l)})+\frac{k\gamma}{2w\mu}\left[\{wR_i(Q)+w'R_i(Q')+\{wC_j(Q)+w'C_j(Q')\}\right]$$

$$+\frac{\gamma^2}{2w\mu}\left[\sum_{i=1}^{u} m_{ij}\{wR_i(Q)+w'R_i(Q')\}+\sum_{j=1}^{u} m_{ij}\{wC_j(Q)+w'C_j(Q')\}\right] \quad \dots \text{(3.24)}$$

where,

$$\mu = k^2-(r-\lambda)(p-q)^2\gamma^2. \qquad \dots \text{(3.25)}$$

By putting $w' = 0$ in (3.24) and (3.25) we can obtain the intra-block estimate of the treatment effect, namely,

$$t_{ij}^{(l)} = \tfrac{1}{2} Q_{ij}^{(l)}+\frac{k\{R_i(Q)+C_j(Q)\}+\sum_{i=1}^{u} m_{ij}R_i(Q)+\sum_{j=1}^{u} m_{ij}C_j(Q)}{2(k^2-(r-\lambda)(p-q)^2)} . \qquad \dots \text{(3.26)}$$

4. Variances of estimated treatment differences

1. We are interested in comparing the effects, as estimated by (3.24), of any two of the v treatments. There are seven types of treatment pairs each having a different expression for the variance of a difference between the two members of the pair. They are separately considered below.

(1) For two treatments belonging to the same cell of the $u \times u$ matrix, the variance of the difference of their effects is $1/w$.

(2) For two treatments belonging to different cells of the same row or column of the $u \times u$ matrix, the variance of the difference in effects is $\frac{1}{w}\left(1+\frac{k\gamma}{\mu}\right)$.

(3) Two treatments not belonging to the same row or column of the $u \times u$ matrix fall into five sub-classes. Let us denote the cells to which these treatments belong by (gh) and (ij) where $g \neq i$ and $h \neq j$. We know that there are two types of cells, namely, those having p treatments and the others having q treatments. If we take the quadrant formed by the four cells :

(gh)	(gj)
(ih)	(ij)

it is clear that there are 16 types of quadrants. They fall into five groups according as the value of the cross difference

$$d = m_{gh} + m_{ij} - m_{gj} - m_{ih} \qquad \ldots \quad (4.1)$$

is either 0, $(p-q)$, $(q-p)$, $2(p-q)$ or $2(q-p)$.

(i) $d = 0$. There are six quadrants in this group, namely,

	h j	h j	h j	h j	h j	h j
g	p p	q q	p p	p q	q p	q q
i	p p	q q	q q	p q	q p	p p

Variance of the difference between the effects of a treatment in cell (gh) and in cell (ij) is $\frac{1}{w}\left(1+\frac{2k\gamma}{\mu}\right)$.

(ii) $d = (p-q)$. There are four quadrants in this group, namely

	h j			
g	p p	p q	p q	q q
i	q p	p p	q q	q p

Variance of the difference between the effects of a treatment in cell (gh) and in cell (ij) is $\frac{1}{w}\left[1+\frac{2k\gamma+(p-q)\gamma^2}{\mu}\right]$.

(iii) $d = (q-p)$. There are four quadrants in this group, namely,

q q	q p	q p	p p
p q	q q	p p	p q

Variance in this case is $\frac{1}{w}\left[1+\frac{2k\gamma+(q-p)\gamma^2}{\mu}\right]$.

(iv) $d = 2(p-q)$. There is only one quadrant in this group, namely,

$$\overline{\begin{array}{cc} p & q \\ \hline q & p \end{array}}$$

Variance is $\dfrac{1}{w}\left[\ 1+\dfrac{2k\gamma+2(p-q)\gamma^2}{\mu}\right].$

(v) $d = 2(q-p)$. There is only one quadrant in this group, namely,

$$\overline{\begin{array}{cc} q & p \\ \hline p & q \end{array}}$$

Variance is $\dfrac{1}{w}\left[\ 1+\dfrac{2k\gamma+2(q-p)\gamma^2}{\mu}\ \right].$

The variances in the cases (i) to (v) may be denoted by the common expression

$$\frac{1}{w}\left[\ 1+\frac{2k\gamma+d\gamma^2}{\mu}\ \right]$$

where $d = 0,\ (p-q),\ (q-p),\ 2(p-q),\ 2(q-p)$ respectively.

2.   Of the total number of $\dfrac{1}{2}\ v(v-1)$ differences between pairs of treatments, the number belonging to type (1) is :

$$\frac{1}{2}\ u\{rp(p-1)+(u-r)q(q-1)\}.$$

The number belonging to type (2) is :

$$u\{rp(k-p)+(u-r)q(k-q)\}.$$

·The number belonging to type (3), sub-types (i) to (v) combined is :

$$\frac{1}{2}\ u\{rp(uk-2k+p)+(u-r)q(uk-2k+q)\}.$$

It is not easy in the general case to split up the last number into those belonging to each of the sub-types (i) to (v) and hence to calculate the mean variance of all comparisons.   For particular designs the number of comparisons of each of the sub-types (i) to (v) can be determined and the mean variance calculated.

3.   *Special cases.* (a) When $u = 3$, r becomes 2 and the values of m_{ij} in the final scheme for the design : $v = 3k$, $k = 2p+q$ are distributed in the form

$$\begin{array}{ccc} p & p & q \\ q & p & p \\ p & q & p \end{array}$$

In this case quadrants with $d = 0$ do not occur. Hence the number of different types of comparisons reduces to six.

Example 1 of Section 2 is a particular case when $p = 2, q = 1$.

(b) When $p > 1$, $q = 0$, comparisons of the sub-types (iii) and (v) do not occur. Hence the number of different types of comparisons reduces to five.

(c) When $p = 0$, $q > 1$, comparisons of the sub-types (ii) and (iv) do not occur. Hence, as in (b), the number of different types of comparisons reduces to five.

(d) When $p = 1, q = 0$ or $p = 0, q = 1$ there are only four types of comparisons. The corresponding variances are

$$\frac{1}{w}\left\{1 + \frac{k\gamma}{\mu}\right\}, \quad \frac{1}{w}\left[1 + \frac{2k\gamma + d\gamma^2}{\mu}\right] \quad (d = 0, 1, 2)$$

where $k = r$ or $u - r$ and $\mu = k^2 - (r - \lambda)\gamma^2$.

(e) If, in (b), $u = r + 1$ so that $\lambda = r - 1$, we have the extended rectangular lattice : $v = u(u - 1)p$, $k = (u - 1)p$ having five variances for treatment comparisons.

(f) If, in (d), $u = r + 1$ so that $\lambda = r - 1$, we have the simple rectangular lattice : $v = u(u - 1)$, $k = (u - 1)$. Here, $\mu = k^2 - \gamma^2$ in the expressions for the four variances.

(g) When $p = q$, which is the same as the case $u = r = \lambda$ the sub-types (i) to (v) merge into a single type of comparison and hence there are only three different variances instead of seven. They are :

$$\frac{1}{w}, \quad \frac{1}{w}\left\{1 + \frac{\gamma}{k}\right\}, \quad \frac{1}{w}\left\{1 + \frac{2\gamma}{k}\right\}.$$

The design is now the extended square lattice : $v = u^2p$, $k = up$.

When $p = q = 1$, there will be only two different variances, namely,

$\frac{1}{w}\left\{1 + \frac{\gamma}{k}\right\}$ and $\frac{1}{w}\left\{1 + \frac{2\gamma}{k}\right\}$. The design becomes the simple square lattice : $v = u^2$, $k = u$.

5. ESTIMATION OF w AND w'

1. The values of w and w' entering into the estimates of treatment effects given by (3.24) and in the variances of the differences of these estimates calculated in Section 4 are not known a priori and have to be estimated from the data. This is done with the aid of the analysis of variance given in Table 1.

2. The following sums of squares must be calculated first

$$\text{Total S.S.} = \sum_{i=1}^{u} \sum_{j=1}^{u} \sum_{l=1}^{m_{ij}} [\{x_{ij}^{(l)^2}\} + \{y_{ij}^{(l)^2}\}] - \frac{\{T(x) + T(y)\}^2}{2v}. \qquad \dots \text{(5.1)}$$

$$\text{Replication S.S.} = \frac{\{T(x) - T(y)\}^2}{2v}. \qquad \dots \text{(5.2)}$$

17

Blocks S.S. (unadjusted) $= \frac{1}{k}\left[\sum\limits_{i=1}^{u}\{R_i(x)\}^2 + \sum\limits_{j=1}^{u}\{C_j(y)\}^2\right] - \frac{1}{v}[\{T(x)\}^2 + \{T(y)\}^2].$ ... (5.3)

Treatments S.S. (unadjusted) $= \frac{1}{2}\sum\limits_{i=1}^{u}\sum\limits_{j=1}^{u}\sum\limits_{l=1}^{m_{ij}}\{x_{ij}^{(l)}+y_{ij}^{(l)}\}^2 - \frac{\{T(x)+T(y)\}^2}{2v}.$ ... (5.4)

3. The treatment sum of squares (adjusted) is calculated using the formula

$$\sum\limits_{i=1}^{u}\sum\limits_{j=1}^{u}\sum\limits_{l=1}^{m_{ij}} t_{ij}^{(l)}.Q_{ij}^{(l)} \qquad\qquad ... (5.5)$$

where $t_{ij}^{(l)}$ is the *intra-block estimate* given in 3.26.

TABLE 1. ANALYSIS OF VARIANCE

source of variation	degrees of freedom	sum of squares	mean square
replication	1	see (5.2)	
blocks within replications (unadjusted)	$2(u-1)$	see (5.3)	
treatments (adjusted)	$(v-1)$	see (5.5)	E_t
intra-block error	$v-2u+1$	by subtraction	E_e
total	$2v-1$	see (5.1)	
blocks within replications (adjusted)	$2(u-1)$	(5.3)+(5.5)—(5.4)	E_b

4. The Blocks Sum of Squares within replications (adjusted) is obtained by using the relation

Block (adj.)+Treatment (unadj.) = Block (unadj.)+Treatment(adj.).

5. A test of significance of the differences among the treatment effects (intra-block) can be performed by the F-test where

$$F = E_t/E_e$$

with degrees of freedom $(v\text{-}1)$ and $(v-2u+1)$.

6. Using formulae given by Nair (1944) the estimates of w and w' are as follows :

$$w' = \frac{1}{2E_b-E_e} ; \quad w = \frac{1}{E_e} \qquad\qquad ... (5.6)$$

Hence,

$$\frac{w'}{w} = \frac{E_e}{2E_b-E_e} \qquad\qquad ... (5.7)$$

and

$$\gamma = \frac{E_b-E_e}{E_b} \qquad\qquad ... (5.8)$$

RESOLVABLE INCOMPLETE BLOCK DESIGNS WITH TWO REPLICATIONS

6. LIST OF DESIGNS

1. Let us list all the designs for which $k \leqslant 10$.

no.	u	r	λ	p	q	k	v	
1	2	2	2	1	0	2	4	S.S.L.*
2	2	2	2	2	0	4	8	E.S.L.*
3	2	2	2	3	0	▊	12	E.S.L.
4	2	2	2	4	0	8	16	E.S.L.
5	2	2	2	5	0	10	20	E.S.L.
6	3	2	1	1	0	▊	6	S.R.L.*
7	3	2	1	1	1	3	9	S.S.L.
8	3	2	1	1	2	4	12	
9	3	2	1	1	3	5	15	
10	3	2	1	1	4	6	18	
11	3	2	1	1	5	7	21	
12	3	2	1	1	6	8	24	
13	3	2	1	1	7	9	27	
14	3	2	1	1	8	10	30	
15	3	2	1	2	0	4	12	E.R.L.*
16	3	2	1	2	1	5	15	
17	3	2	1	2	2	6	18	E.S.L.
18	3	2	1	2	3	7	21	
19	3	2	1	2	4	8	24	
20	3	2	1	2	5	9	27	
21	3	2	1	2	6	10	30	
22	3	2	1	3	0	▊	18	E.R.L.
23	3	2	1	3	1	7	21	
24	3	2	1	3	2	8	24	
25	3	2	1	3	3	9	27	E.S.L.
26	3	2	1	3	4	10	30	
27	3	2	1	4	0	8	24	E.R.L.
28	3	2	1	4	1	9	27	
29	3	2	1	4	2	10	30	
30	3	2	1	5	0	10	30	E.R.L.

See footnote on the last page of the list

19

no.	u	r	λ	p	q	k	v	
31	4	3	2	1	0	3	12	S.R.L.
32	4	3	2	1	1	4	16	S.S.L.
33	4	3	2	1	2	5	20	
34	4	3	2	1	3	6	24	
35	4	3	2	1	4	7	28	
36	4	3	2	1	5	8	32	
37	4	3	2	1	6	9	36	
38	4	3	2	1	7	10	40	
39	4	3	2	2	0	6	24	E.R.L.
40	4	3	2	2	1	7	28	
41	4	3	2	2	2	8	32	E.S.L.
42	4	3	2	2	3	9	36	
43	4	3	2	2	4	10	40	
44	4	3	2	3	0	9	36	E.R.L.
45	4	3	2	3	1	10	40	
46	5	4	3	1	0	4	20	S.R.L.
47	5	4	3	1	1	5	25	S.S.L.
48	5	4	3	1	2	6	30	
49	5	4	3	1	3	7	35	
50	5	4	3	1	4	8	40	
51	5	4	3	1	5	9	45	
52	5	4	3	1	6	10	50	
53	5	4	3	2	0	8	40	E.R.L.
54	5	4	3	2	1	9	45	
55	5	4	3	2	2	10	50	E.S.L.
56	6	5	4	1	0	5	30	S.R.L.
57	6	5	4	1	1	6	36	S.S.L.
58	6	5	4	1	2	7	42	
59	6	5	4	1	3	8	48	
60	6	5	4	1	4	9	54	
61	6	5	4	1	5	10	60	
62	6	5	4	2	0	10	60	E.R.L.
63	7	6	5	1	0	6	42	S.R.L.
64	7	6	5	1	1	7	49	S.S.L.
65	7	6	5	1	2	8	56	
66	7	6	5	1	3	9	63	
67	7	6	5	1	4	10	70	
68	8	7	6	1	0	7	56	S.R.L.
69	8	7	6	1	1	8	64	S.S.L.

RESOLVABLE INCOMPLETE BLOCK DESIGNS WITH TWO REPLICATIONS

no.	u	r	λ	p	q	k	v	
70	8	7	6	1	2	9	72	
71	8	7	6	1	3	10	80	
72	9	8	7	1	0	8	72	S.R.L.
73	9	8	7	1	1	9	81	S.S.L.
74	9	8	7	1	2	10	90	
75	10	9	8	1	0	9	90	S.R.L.
76	10	9	8	1	1	10	100	S.S.L.
77	11	10	9	1	0	10	110	S.R.L.
78	7	3	1	1	0	3	21	
79	7	3	1	0	1	4	28	
80	7	3	1	2	0	6	42	
81	7	3	1	0	2	8	56	
82	7	3	1	2	1	10	70	
83	13	4	1	1	0	4	52	
84	13	4	1	0	1	9	117	
85	13	4	1	2	0	8	104	
86	21	5	1	1	0	5	105	
87	21	5	1	2	0	10	210	
88	31	6	1	1	0	6	186	
89	57	8	1	1	0	8	456	
90	73	9	1	1	0	9	657	
91	91	10	1	1	0	10	910	
92	11	5	2	1	0	5	55	
93	11	5	2	0	1	6	66	
94	11	5	2	2	0	10	110	
95	16	6	2	1	0	6	96	
96	16	6	2	0	1	10	160	
97	37	9	2	1	0	9	333	
98	15	7	3	1	0	7	105	
99	15	7	3	0	1	8	120	
100	25	9	3	1	0	9	225	
101	31	10	3	1	0	10	310	
102	19	9	4	1	0	9	171	
103	19	9	4	0	1	10	190	

S.S.L. = Simple Square Lattice [$v=u^2$, $k=u$, $r=2$, $b=2u$]

S.R.L. = Simple Rectangular Lattice [$v=u(u-1)$, $k=(u-1)$, $r=2$, $b=2u$]

E.S.L. = Extended Square Lattice [$v=u^2p$, $k=up$, $r=2$, $b=2u$]

E.R.L. = Extended Rectangular Lattice [$v=u(u-1)p$, $k=(u-1)p$, $r=2$, $b=2u$]

We have 103 designs for $k \leqslant 10$ out of which 9 each are simple square lattices and simple rectangular lattices. If arranged according to magnitude of v and k, we have the following scheme

v =	4	6	8	9	12	15	16	18	20	21	24	25	27	28	30	32	35	36	40
k =	2	2	4	3	3	5	4	6	4	3	6	5	9	4	5	8	7	6	8
					4	5	8	6	5	7	6	−	9	7	6	8	−	9	8
					4			6	10	7	8		9	7	10			9	10
					6					7	8		9	−	10			9	10
					−					−	8		−		10			−	10
											8				10				−
											−				10				

v =	42	45	48	49	50	52	54	55	56	60	63	64	66	70	72	80	81
k =	6	9	8	7	10	4	9	5	7	10	9	8	6	10	8	10	9
	6	9			10				8	10				10	9		
	7	−			−				8	−							
	−								−								

v =	90	96	100	104	105	110	117	120	160	171	186	190	210	225	310	333
k =	9	6	10	8	5	10	9	8	10	9	6	10	10	9	10	9
	10				7	10										
	−				−	−										

v =	456	657	910
k =	8	9	10

2. The simple and extended square lattices and the simple and extended rectangular lattices are PBIB designs with two, three, four and five associate classes respectively. The number of different variances to be calculated for treatment differences in these cases is 2, 3, 4 and 5 respectively.

Out of the remaining designs listed above we have found that those designs derived from the symmetrical balanced incomplete block design

$$v = s^2 + s + 1$$
$$r = s + 1$$
$$\lambda = 1$$

and for which $p > 0$ and $q = 0$, so that

$$k = p(s+1)$$
$$v = p(s+1)(s^2+s+1)$$

are PBIB designs with four associate classes.

The parameters are :

$$v = p(s+1)(s^2+s+1) \qquad\qquad r = 2$$
$$k = p(s+1) \qquad\qquad\qquad\quad b = 2(s^2+s+1)$$
$$\lambda_1 = 2 \qquad \lambda_2 = 1 \qquad \lambda_3 = 0 \qquad \lambda_4 = 0$$
$$n_1 = (p-1) \qquad n_2 = 2ps \qquad n_3 = 2ps^2 \qquad n_4 = ps^3$$

$$p^1_{jk} = \begin{pmatrix} (p-2) & 0 & 0 & 0 \\ 0 & 2ps & 0 & 0 \\ 0 & 0 & 2ps^2 & 0 \\ 0 & 0 & 0 & ps^3 \end{pmatrix}$$

$$p^2_{jk} = \begin{pmatrix} 0 & (p-1) & 0 & 0 \\ (p-1) & p(s-1) & ps & 0 \\ 0 & ps & ps(s-1) & ps^2 \\ 0 & 0 & ps^2 & ps^2(s-1) \end{pmatrix}$$

$$p^3_{jk} = \begin{pmatrix} 0 & 0 & (p-1) & 0 \\ 0 & p & p(s-1) & ps \\ (p-1) & p(s-1) & ps & 2ps(s-1) \\ 0 & ps & 2ps(s-1) & ps(s-1)^2 \end{pmatrix}$$

$$p^4_{jk} = \begin{pmatrix} 0 & 0 & 0 & (p-1) \\ 0 & 0 & 2p & 2p(s-1) \\ 0 & 2p & 4p(s-1) & 2p(s-1)^2 \\ (p-1) & 2p(s-1) & 2p(s-1)^2 & p(s-1)(s^2-s+1) \end{pmatrix}$$

If $p = 1$, the design becomes a PBIB design with three associate classes. The parameters are :

$$v = (s+1)(s^2+s+1) \qquad\qquad r = 2$$
$$k = (s+1) \qquad\qquad\qquad\quad b = 2(s^2+s+1)$$

$$\lambda_1 = 1 \qquad \lambda_2 = 0 \qquad \lambda_3 = 0$$
$$n_1 = 2s \qquad n_2 = 2s^2 \qquad n_3 = s^3$$

$$p^1_{jk} = \begin{pmatrix} (s-1) & s & 0 \\ s & s(s-1) & s^2 \\ 0 & s^2 & s^2(s-1) \end{pmatrix} \qquad p^2_{jk} = \begin{pmatrix} 1 & (s-1) & s \\ (s-1) & s & 2s(s-1) \\ s & 2s(s-1) & s(s-1)^2 \end{pmatrix}$$

$$p^3_{jk} = \begin{pmatrix} 0 & 2 & 2(s-1) \\ 2 & 4(s-1) & 2(s-1)^2 \\ 2(s-1) & 2(s-1)^2 & (s-1)(s^2-s+1) \end{pmatrix}$$

23

3. Among the remaining designs it is likely that some are PBIB designs. For instance, design No. 79 is a PBIB design with the following parameters.

$$v = 28 \qquad k = 4 \qquad r = 2 \qquad b = 14$$

$$\lambda_1 = 1 \qquad \lambda_2 = 0 \qquad \lambda_3 = 0 \qquad \lambda_4 = 0$$

$$n_1 = 6 \qquad n_2 = 3 \qquad n_3 = 12 \qquad n_4 = 6$$

$$p_{jk}^1 = \begin{pmatrix} 2 & 1 & 2 & 0 \\ 1 & 0 & 2 & 0 \\ 2 & 2 & 4 & 4 \\ 0 & 0 & 4 & 2 \end{pmatrix} \qquad p_{jk}^2 = \begin{pmatrix} 2 & 0 & 4 & 0 \\ 0 & 0 & 0 & 2 \\ 4 & 0 & 4 & 4 \\ 0 & 2 & 4 & 0 \end{pmatrix}$$

$$p_{jk}^3 = \begin{pmatrix} 1 & 1 & 2 & 2 \\ 1 & 0 & 1 & 1 \\ 2 & 1 & 6 & 2 \\ 2 & 1 & 2 & 1 \end{pmatrix} \qquad p_{jk}^4 = \begin{pmatrix} 0 & 0 & 4 & 2 \\ 0 & 1 & 2 & 0 \\ 4 & 2 & 4 & 2 \\ 2 & 0 & 2 & 1 \end{pmatrix}$$

It follows from this that design No. 81 is a PBIB design with five associate classes.

4. Designs with both p and q unequal and different from zero will not be PBIB designs.

REFERENCES

BOSE, R. C. and NAIR, K. R. (1939) : Partially balanced incomplete block designs. *Sankhyā*, **4**, 337-372.

BOSE, R. C. (1951) : Partially balanced incomplete block designs with two associate classes involving only two replications. *Cal. Stat. Ass. Bull.*, **3**, 120-125.

NAIR, K. R. and RAO, C. R. (1942) : A note on partially balanced incomplete block designs. *Science and Culture*, **7**, 568-569.

NAIR, K. R. (1944) : The recovery of inter-block information in incomplete block designs. *Sankhyā*, **6**, 383-390.

———— (1950) : Partially balanced incomplete block designs involving only two replications. *Cal. Stat. Ass. Bull.*, **3**, 83-86.

———— (1951) : Some two-replicate partially balanced designs. *Cal. Stat. Ass. Bull.*, **3**, 174-176.

YATES, F. (1936) : Incomplete randomized blocks. *Ann. Eugen.*, **7**, 121-140.

Paper received : November, 1961.

ON A METHOD OF CONSTRUCTING SYMMETRICAL BALANCED INCOMPLETE BLOCK DESIGNS

By S. S. SHRIKHANDE

and

N. K. SINGH

Banaras Hindu University

SUMMARY. A new method for the construction of Symmetrical Balanced Incomplete Block designs is indicated which makes use of the existence of certain association schemes useful in Partially Balanced Incomplete Block designs. Some related problems are also considered.

1. INTRODUCTION

A Balanced Incomplete Block Design (BIBD) is an arrangement of v elements (treatments) into b sets (blocks) of $k(<v)$ distinct elements such that every pair of different elements occurs in exactly λ blocks. Then it is easy to see that every element occurs in exactly r of these sets. The numbers v, b, r, k, λ are called the parameters of the design and they satisfy the relations

$$vr = bk, \quad \lambda(v-1) = r(k-1), \quad b \geqslant v. \qquad \ldots \quad (1.1)$$

The last inequality is due to Fisher (1940). A BIBD is called symmetrical if $v = b$ and hence $r = k$. Systematic methods of construction for BIBD have been given by various authors, for example, Bose (1939), Rao (1945, 1946) Sprott (1954), Bose and Shrikhande (1960), Haim Hanani (1961).

Following Bose and Shimamoto (1952) we say that an m associate classes association scheme for v treatments is defined if the following conditions are satisfied.

(a) Any two treatments are either first associates, or second associates... or m-th associates.

(b) Each treatment has n_i i-th associates, $i = 1, 2 \ldots, m$.

(c) Given any two treatments which are i-th associates, the number p^i_{jk} of treatments which are j-th associates of the first and k-th associates of the second is independent of the pair of treatments with which we start. Further, $p^i_{jk} = p^i_{kj}$, for $i, j, k = 1, 2, \ldots, m$.

A Partially Balanced Incomplete Block Design (PBIBD) is an arrangement of v treatments in b blocks such that

(1) each of the v treatments is replicated r times in b blocks of size k and no treatment occurs more than once in any block;

(2) there exists an m ($\geqslant 2$) associate classes association scheme for v treatments as defined above;

(3) any pair of treatments which are i-th associates occur together in exactly λ_i blocks for $i = 1, 2, \ldots, m$, where all the λ_i's are not equal.

The definition of PBIBD as given above is essentially due to Bose and Nair (1939) with a slight modification by Nair and Rao (1942). The numbers $v, b, r, k,$ $\lambda_1, ..., \lambda_m$ are called parameters of the first kind, whereas the numbers $n_1, n_2, ..., n_m$ and p^i_{jk} are called the parameters of the second kind. It is possible for different PBIB designs to have the same association scheme but different parameters of the first kind. The important idea of separating a PBIBD from its association scheme is due to Bose and Shimamoto as mentioned earlier. A crucial role in the existence of of some symmetrical BIB designs is played by certain types of association schemes and PBIB designs are used only in so far as they provide the existence of these association schemes.

2. Construction of symmetrical BIB designs

Consider a m classes association scheme for v treatments with parameters $n_i, p^i_{jk}, i, j, k = 1, 2, ..., m$. The following relations are known to exist

$$\sum_{i=1}^{m} n_i = v-1 \qquad\qquad ... \quad (2.1)$$

$$\sum_{k=1}^{m} p^i_{jk} = \begin{cases} n_j & \text{if } i \neq j \\ n_i-1 & \text{if } i = j. \end{cases} \qquad ... \quad (2.2)$$

We write the parameters p^i_{jk} in the form of m matrices $P_i = (p^i_{jk})$, $i, j, k = 1, 2, ..., m$, which are necessarily symmetrical. In view of (2.2), the sum of the elements in the j-th row of P_i is n_j if $j \neq i$ and n_i-1 if $j = i$.

Now suppose that the matrices $P_1, P_2, ..., P_m$, satisfy the condition that the sum of the elements in the square submatrices in each corresponding to rows and columns numbered $i_1, i_2, ..., i_p$, $p < m$ is a constant, say, λ. Define two treatments to be α-th associates if they are either i_1-th, i_2-th, ..., i_p-th associates; otherwise define them as β-th associates. Obviously, then n_α and n_β which are the numbers of α-th and β-th associates of a treatment are given by,

$$n_\alpha = n_{i_1} + n_{i_2} + ... + n_{i_p}$$
$$n_\beta = (v-1) - n_\alpha. \qquad\qquad ... \quad (2.3)$$

If two treatments θ and ϕ are α-th associates the number $p^\alpha_{\alpha\alpha}$ of treatments which are simultaneously the α-th associates of θ and ϕ is the number of treatments common to α-th associates of θ and ϕ. If θ and ϕ are i_1-th associates, this number is easily seen to be the sum of the elements in the square submatrix of P_{i_1} in rows and columns corresponding to $i_1, i_2, ..., i_p$ and hence is equal to λ. Similarly this number is λ if θ and ϕ are i_2-th, i_3-th, ... or i_p-th associates. Hence $p^\alpha_{\alpha\alpha} = \lambda$ irrespective of what two treatments we start with so long as they are α-th associates. A similar argument shows that $p^\beta_{\alpha\alpha} = \lambda$, for any two treatments which are β-th associates. It now follows from Bose and Clatworthy (1955), that we get a two associate classes association scheme with parameters n_α, n_β and $p_{jk}, i, j, k = \alpha, \beta$, such that

$$P_\alpha = \begin{pmatrix} \lambda & n_\alpha-1-\lambda \\ & n_\beta-n_\alpha+\lambda+1 \end{pmatrix}, \quad P_\beta = \begin{pmatrix} \lambda & n_\alpha-\lambda \\ & n-n_\alpha+\lambda-1 \end{pmatrix} \qquad ... \quad (2.4)$$

We can, therefore, state the following theorem.

Theorem 1 : *If we have an m associate classes association scheme with parameters n_i and matrices P_i, $i = 1, 2, ..., m$, such that the sum of the elements in these matrices standing in rows and columns numbered $i_1, i_2, ..., i_p$ is a constant λ, then we can form a two associate classes association scheme with parameters n_α, n_β and matrices P_α and P_β given by* (2.3) *and* (2.4).

Theorem 2 : *If there exists a two classes association scheme with parameters n_1, n_2 and matrices P_1 and P_2 such that either*

$$\text{(i)} \quad p^1_{11} = p^2_{11} = \lambda \qquad or \qquad \text{(ii)} \quad p^1_{22} = p^2_{22} = \lambda$$

then we can construct a symmetrical BIBD with parameters v, $r = n_1$, $\lambda = \lambda$ or v, $r = n_2$, $\lambda = \lambda$ according as condition (i) *or* (ii) *is satisfied.*

Proof : Suppose condition (i) is satisfied. Consider the v blocks formed by first-associates of each of the v treatments. In virtue of (i), any two of these blocks intersect in exactly λ treatments. We thus have v blocks of size $k = n_1$ involving v treatments. Since each treatment has exactly n_1 first associates, each treatment occurs in exactly $r = n_1$ of these blocks. It then follows from results of Chowla and Ryser (1950), that we have a symmetric BIBD with parameters v, $r = n_1$, $\lambda = \lambda$. In the same manner if conditionn (ii) is satisfied we get a symmetrical BIBD with parameters v, $r = n_2$, $\lambda = \lambda$.

Corollary 1 : *Suppose there exists a BIBD with $\lambda = 1$ and $r = 2k+1$, then we can construct a symmetrical BIBD with parameters*

$$v = 4k^2 - 1, \quad r = 2k^2, \quad \lambda = k^2 \qquad \qquad \text{... .(2.5)}$$

or equivalently a Hadamard matrix of order $4k^2$.

Proof : Using the conditions (1.1), it is easy to verify that the parameters of the given BIBD are

$$v = k(2k-1), \quad b = 4k^2-1, \quad r = 2k+1, \quad k = k, \quad \lambda = 1.$$

As shown by Shrikhande (1952), the dual of this design is a PBIBD with two associate classes with parameters

$$v = 4k^2-1, n_1 = 2k^2, \quad p^1_{11} = p^2_{11} = k^2.$$

The result now follows from Theorem 2, coupled with the equivalence of a BIBD with parameters (2.5) and a Hadamard matrix of order $4k^2$ (Todd, 1933).

Example 1 : BIB designs of the above corollary are known to exist for $k = 2, 3, 4, 5, 6, 7$ from Fisher and Yates's Tables (1949) and Rao (1961). This guarantees the existence of BIB designs with parameters (2.5) corresponding to $v = 15$, 35, 63, 99, 143 and 195.

Corollary 2 : *Existence of $k-2$ mutually orthogonal latin squares* (m.o.l.s.) *of order $2k$ implies the existence of symmetrical BIBD with parameters*

$$v = 4k^2, \quad r = k(2k-1), \quad \lambda = k(k-1) \qquad \qquad \text{... (2.6)}$$

Proof : From Bose and Shimamoto (1952) the existence of $k-2$ m.o.l.s. of order $2k$ implies the existence of a two associate classes association scheme with $v = 4k^2, n_1 = k(2k-1), \lambda = k(k-1)$. The result now follows from the above theorem.

Example 2 : Taking $k = 2, 3, 4$ the conditions of the corollary are obviously satisfied. For $k = 6$, the existence of 5 m.o.l.s. is guaranteed by Bose, Chakravarti and Knuth (1960). Hence a BIBD with parameters (2.6) exists for those values of k.

It is conjectured by Dulmage, Johnson and Mendelsohn (1961) that $2p-1$ m.o.l.s. of order $4p$ exist if p is a prime. If this conjecture is true, it will guarantee the existence of the symmetrical BIBD with parameters

$$v = 16p^2, \ r = 2p(4p-1), \ \lambda = 2p(2p-1)$$

for every prime p.

Corollary 3 : *If* $s = 2^m$, *then the symmetrical BIBD with parameters*

$$v = s^2(s+2), \ r = s(s+1), \ \lambda = s$$

exists.

Proof : If $s = 2^m$, from the results of Ray Chaudhuri (1959) it is known that a PBIBD with two associate classes exists with the following parameters

$$v = s^3, b = s^2(s+2), \ r = s+2, k = s, \ \lambda_1 = 1, \ \lambda_2 = 0$$
$$n_1 = (s+2)(s-1), p_{11}^1 = s-2, p_{11}^2 = s+2.$$

This design belongs to the series given by Bose and Clatworthy (1955) with $t = 1$ in their notation and its dual is another PBIBD with two associate classes and parameters

$$v = s^2(s+2), \ b = s^3, \ \dot{r} = s, \ k = s+2, \ \lambda_1 = 1, \ \lambda_2 = 0,$$
$$n_1 = s(s+1), p_{11}^1 = \dot{p}_{11}^2 = s.$$

The result now follows from the above theorem.

Example 3 : Taking $s = 2, 4, 8$ we get designs with parameters (16, 6, 2) (96, 20, 4) and (640, 72, 8). The design (96, 20, 4) is mentioned as an unsolved case for difference sets in Marshall Hall (1956).

Example 4 : Consider the association scheme for the design sl. 14 given by Bose, Clatworthy, Shrikhande (1954) having the parameters $v = 45$, $n_1 = 12$, $p_{11}^1 = p_{11}^2 = 3$. This gives the symmetrical design with parameters (45, 12, 3) which is left as an unsolved case by Rao (1961).

In all the above examples, we have made use of Theorem 2 only. Theorem 1 can also be used in a similar manner, if we can construct for any given value of v at least one PBIBD with m-associate classes having the required properties, where the number m is any integer $\geqslant 2$. It should be pointed that certain association schemes with two associate classes can be proved impossible by using results of Connor and Clatworthy (1954), in which case because of Theorem 1 an association scheme with $m \geqslant 3$ associate classes useful for our purpose is automatically ruled out. However in a case where two associate classes association scheme is not proved impossible, it may often be easier to construct a PBIB design with $m \geqslant 3$ associate calsses rather than with two associate classes.

3. Impossibility of certain association schemes

Connor and Clatworthy (1954) have given the following results on the impossibility of an association scheme with two associate classes. Defining

$$\Delta = y^2 + 2\beta + 1$$
$$y = p^2_{12} - p^1_{12}$$
$$\beta = p^2_{12} + p^1_{12} \qquad \qquad \ldots \ (3.1)$$
$$2\eta\sqrt{\Delta} = (v-1)(1-y) - 2n_1$$
$$2\alpha_1 = (v-1) + 2\eta$$
$$2\alpha_2 = (v-1) - 2\eta$$

they state the following theorems.

Theorem A : *If in a PBIBD with two associate classes v is odd and Δ is not an integral square, then it is necessary that*

$$\text{(i)} \quad p^1_{12} = p^2_{12} = p^2_{11} = \frac{v-1}{4} = t,$$

$$\text{(ii)} \quad n_1 = n_2 = \alpha_1 = \alpha_2 = 2t,$$

and (iii) $\quad v = \Delta = 4t+1.$

Theorem B : *If in a PBIBD with two associate classes v is odd and Δ is an integral square, then η must be an integer.*

Theorem C : *If in a PBIBD with two associate classes v is even, then it is necessary that*

$$\text{(i)} \quad \Delta \text{ must be an integral square,}$$

and (ii) $\quad 2\eta$ *must be an odd (positive or negative) integer.*

Consider a two classes association scheme given by one of the following where $4t$ is not an integral square.

$$v = 4t-1, \qquad n_1 = 2t-1, \qquad p^1_{11} = p^2_{11} = t-1 \qquad \qquad \ldots \ (3.2)$$
$$v = 4t-1, \qquad n_1 = 2t, \qquad p^1_{11} = p^2_{11} = t \qquad \qquad \ldots \ (3.3)$$
$$v = 4t-1, \qquad n_2 = 2t-1, \qquad p^1_{22} = p^2_{22} = t-1 \qquad \qquad \ldots \ (3.4)$$
$$v = 4t-1, \qquad n_2 = 2t, \qquad p^1_{22} = p^2_{22} = t \qquad \qquad \ldots \ (3.5)$$

In all these cases it is easy to verify that $\Delta = 4t \neq$ an integral square. Hence from Theorem A all these association schemes are impossible. Thus a Hadamard type symmetric BIBD with parameters $(4t-1, 2t-1, t-1)$ or $(4t-1, 2t, t)$ cannot be obtained by our method if $4t \neq$ an integral square. As is well known such designs do exist for a large number of values of t and it is actually conjectured that they exist for all t. Bose and Shrikhande (1959) give a complete bibliography of the existence problem of such designs.

We now give an example of an association scheme which satisfies all the conditions of Connor and Clatworthy and yet is impossible. Consider a two classes association scheme with

$$v = 4t+1, \quad n_1 = n_2 = 2t$$

$$P_1 = \begin{pmatrix} t-1 & t \\ t & t \end{pmatrix}, \quad P_2 = \begin{pmatrix} t & t \\ t & t-1 \end{pmatrix} \qquad \ldots \text{(3.6)}$$

It is easy to verify that we get $\Delta = 4t+1$, and $\eta = 0$. If $4t+1 \neq$ an integral square, then all the conditions of Theorem A are satisfied, and if $4t+1 =$ an integral square, then again all the conditions of Theorem B are satisfied.

Now define a square matrix $M = (m_{ij})$ of order $4t+1$ by

$$m_{ii} = 0,$$

$m_{ij} = 1$ if treatments i and j are first associates,

$$= -1 \text{ otherwise.} \qquad \ldots \text{(3.7)}$$

Then utilising (3.6) it is easy to verify that

$$MM' = (4t+1)I_{4t+1} - G_{4t+1} \qquad \ldots \text{(3.8)}$$

where M' is the transpose of M and I_n and G_n respectively denote the identity matrix of order n and the square matrix of order n with all elements equal to 1. The matrix MM' is obviously singular.

Define the square matrix N of order $4t+2$ by

$$N = \begin{pmatrix} 0 & J \\ J' & M \end{pmatrix}, \qquad \ldots \text{(3.9)}$$

where J is a row vector with $4t+1$ components all equal to 1.

Then
$$NN' = (4t+1)I_{4t+2} \qquad \ldots \text{(3.10)}$$

Utilising the Minkowski-Hasse Theorem (Hasse, 1923) as given by Bruck and Ryser (1949) we obtain

$$c_p(NN') = (-1, \, 4t+1)_p \qquad \ldots \text{(3.11)}$$

where p is an odd prime. If p is a factor of the square free part of $4t+1$ and is congruent to 3(mod 4), then

$$c_p(NN') = -1. \qquad \ldots \text{(3.12)}$$

Since NN' is rationally congruent to I_{4t+2}

$$c_p(I_{4t+2}) = c_p(NN') = 1. \qquad \ldots \text{(3.13)}$$

Hence we get a contradiction. We may, therefore, state

Theorem 3 : *A two associate classes association scheme given by (3.6) is non-existent if the square free part of $4t+1$ contains a prime congruent to 3 (mod 4).*

Thus for $t = 5, 8, 14$ the association scheme (3.6) is impossible.

In the proof of the above theorem we have already established that the existence of the association scheme given by (3.6) implies the existence of a square matrix N of order $4t+2$ with components 0, 1, -1 such that $NN' = (4t+1)I_{4t+2}$. We now prove the converse result. Given such a matrix N, it is obvious that 0 occurs once and only once in every row and column and hence by suitable interchanges of columns we can bring zeroes along the main diagonal. Again interchanging 1 and -1 in all the columns except the first, if necessary, we can make all the elements of the first row 1 excepting for the single 0 in the initial position. Similarly we can change all the elements of the first column excepting the single 0 into 1. Hence without loss of generality we can write

$$N = \begin{pmatrix} 0 & J \\ J' & M \end{pmatrix}$$

where J and J' are row and column vectors with all components 1. It is easy to see that

$$MM' = (4t+1)I_{4t+1} - G_{4t+1}$$

and that each row and column of M contains 1 and -1 exactly $2t$ times. Since the scalar product of any two rows of M is -1, and 0 occurs only along the main diagonal, it is easy to verify that M is a symmetric matrix. For suppose without loss of generality that the first two rows of M are given by

		2t−1 times		2t times		
0	1	1········1	1··············1	−1	−1 −1	−1
−1	0	1········1	−1··············−1	1···········1	−1··········−1	
		α times	2t−1−α times	2t−α times	α times	

Then their scalar product is $4\alpha - 4t + 1$ which can never be equal to -1. Hence the element in position $(2, 1)$ must be the same as the element in position $(1, 2)$, and hence the matrix M must be symmetrical. Again it can be easily verified that if the element (i, j) $i \neq j$ is 1, then the rows i and j contain the ordered pairs $(1, 1)$ $(1, -1)$, $(-1, 1)$ and $(-1, -1)$ exactly $t-1$, t, t, t times respectively. Similarly if the element (i, j) $= -1$, then these pairs occur t, $t, t, t-1$ times respectively. Identifying the rows and columns of M with a set of $4t+1$ treatments and taking $1(-1)$ in position (i, j) to imply that treatment j is first (second) associate of treatment i, we obviously have a two classes association scheme with parameters given by (3.6).

In conclusion it may be mentioned that the matrix N has been used by Raghavarao (1960) in certain weighing designs, and the nonexistence of N has been proved by him under identical conditions.

REFERENCES

Bose, R. C. (1939): On the construction of balanced incomplete block designs. *Ann. Eugen.* **9**, 353-399.

Bose, R. C. and Nair, K. R., (1939): Partially balanced incomplete block designs. *Sankhya*, **4**, 337-372.

Bose, R. C. and Shimamoto, T. (1952): Classification and analysis of partially balanced incomplete block designs with two associate classes. *J. Amer. Stat. Ass.*, **47**, 151-184.

Bose, R. C., Clatworthy, W. H. and Shrikhande, S. S. (1954): Tables of partially balanced designs with two associate classes. North Carolina Agricultural Experiment Station, *Technical Bulletin* No. 107.

Bose, R. C. and Clatworthy, W. H. (1955): Some classes of partially balanced designs. *Ann. Math. Stat.*, **26**, 212-232.

Bose, R. C. and Shrikhande, S. S. (1959): A note on a result in the theory of code construction. *Information and Control*. **2**, 183-194.

————(1960): On the composition of balanced incomplete block designs. *Can. J. Math.* **12**, 177-188.

Bose, R. C., Chakravarti, I. M. and Knuth, D. E., (1960): On methods of constructing sets of mutually orthogonal Latin squares using a computer—I. *Technometrics*, **2**, 507-516.

Bruck, R. H. and Ryser, H. J. (1949): Nonexistence of certain finite projective planes. *Can. J. Math.* **1**, 88-93.

Chowla, S. and Ryser, H. J. (1950): Combinatorial problems. *Can. J. Math.* **2**, 93-99.

Connor, W. S. and Clatworthy, W. H. (1954): Some theorems for partially balanced designs. *Ann. Math. Stat.* **25**, 100-112.

Fisher, R. A., (1940): An examination of the different possible solutions of a problem in incomplete blocks. *Ann. Eugen. London*, **10**, 52-75.

Fisher, R. A. and Yates, F., (1949): *Statistical Tables*. Hafner Publishing Company, New York.

Hall, Marshall Jr. (1956): A survey of difference sets. *Proc. Amer. Math. Soc.* **7**, 975-986.

Hanani, H. (1961): The existence and construction of balanced incomplete block designs. *Ann. Math. Stat.* **32**, 361-386.

Hasse, H. (1923): Uber die Aquivalenz quadratischer formen in körper rationalen zahlen. *J. Reine Agnew. Math.* **152**, 205-224.

Johnson,, D. M., Dulmage, A. L. and Mendllsohn, N. S. (1961): Orthomorphisms of groups and orthogonal Latin Squares I. *Can. J. Math* **13**, 356-372.

Nair, K. R. and Rao, C. R. (1942): A note on partially balanced incomplete block designs. *Science and Culture*. **7**, 568-569.

Raghavarao, D. (1960) Some aspects of weighing designs. *Ann. Math. Stat.* **31**, 878-884.

Rao, C.R. (1945): Finite geometries and certain derived results in theory of numbers. *Proc. Nat. Inst. Sc.* **11**, 136-149.

———— (1946): Difference sets and combinatorial arrangements derivable from finite geometries. *ibid*, **12**, 123-135.

———— (1961): A study of BIB designs with replications 11 to 15. *Sankhyā*. Series A, **23**, 117-127.

Ray Chaudhuri, D. K. (1959): On the application of the geometry of quadrics to the construction of partially balanced incomplete block designs and error correcting binary codes. Ph.D. Thesis, University of North Carolina, Chapel Hill, U.S.A.

Shrikhande, S. S. (1952): On the dual of some balanced incomplete block designs. *Biometrics*. **8**, 66-72.

Sprott, D. A. (1954): Note on balanced incomplete block designs. *Can. J. Math.* **6**, 34-46.

Todd, J. A. (1933): A combinatorial problem. *J. Math. and Phys.* **12**, 321-333.

Paper received : November, 1961.

ON A TWO-PARAMETER FAMILY OF BALANCED INCOMPLETE BLOCK DESIGNS

By S. S. SHRIKHANDE

Banaras Hindu University

SUMMARY. A general two-parameter family of Balanced Incomplete Block (BIB) designs is derived and a method of constructing them is given. Connection of a sub-family of these designs with Hadamard matrices and certain optimum minimum distance codes is also established.

1. INTRODUCTION

In a recent paper P. K. Menon (1962) has given a one-parameter family of difference sets for symmetric BIB designs. In this paper we generalise his results to derive a two parameter family of BIB designs such that any two members of the family can be used to obtain a new member of the same family. My thanks are due to Menon for having shown me a copy of his manuscript which led to the present results.

2. NOTATION AND PRELIMINARY RESULTS

A BIB design with parameters v, b, r, k, λ is an arrangement of v objects or treatments in b sets or blocks such that (i) every block contains $k < v$ different treatments, and (ii) every pair of distinct treatments occurs in exactly λ blocks. Then it is easy to see that every treatment occurs in exactly r blocks and the parameters satisfy the relations

$$\lambda(v-1) = r(k-1), \quad bk = vr, \quad b \geqslant v. \qquad \ldots \quad (2.1)$$

The last inequality is due to Fisher (1940). A BIB design is called symmetrical if $b = v$ and hence $r = k$.

Let $N = (n_{ij})$ be the incidence matrix of order (v, b) of a BIB design with above parameters, i.e.,

$$n_{ij} = 1 \text{ if treatment } i \text{ occurs in block } j, \qquad \ldots \quad (2.2)$$
$$= 0 \text{ otherwise.}$$

Then it is obvious that every row and column of N contains 1 in r and k places respectively and that any two rows of N contain the pair $(1, 1)$ in exactly λ columns. Conversely any $(0, 1)$ matrix having these properties can be regarded as the incidence matrix of a BIB design.

If N is the incidence matrix of the above BIB design, then by interchanging 0 and 1, we get

$$N^* = J - N \qquad \ldots \quad (2.3)$$

where J is the matrix of order (v, b) with all elements 1. It is easy to verify that N^* is the incidence matrix of the complementary BIB design with parameters $v, b, b-r, v-k, b-2r+\lambda$.

If $A = (a_{ij})$ and $B = (b_{ij})$ are two matrices of order (m, n) and (p, q) respectively, then the Kronecker product of A and B, denoted by $A \times B$ is defined by

$$C = A \times B = (a_{ij} B), \quad i = 1, 2, ..., m; \quad j = 1, 2, ..., n. \qquad \ldots \text{(2.4)}$$

The matrix C is of order (mp, nq). Kronecker product of matrices has been used by Vartak (1955) in the construction of Partially Balanced Incomplete Block designs.

A Hadamard matrix H_n is a $(1, -1)$ square matrix of order n such that $H_n H'_n = n I_n$ where I_n is the identity matrix of order n. It is known that H_n can exist only when n is either 2 or a multiple of 4. A complete summary of the status of the existence problem is given by Bose and Shrikhande (1959) and it is conjectured that a Hadamard matrix of every possible order exists.

3. Composition of BIB designs

We give here the generalisation of the method given by Menon (1962).

Theorem 1 : *If D_i is a BIB design with parameters $v_i, b_i, r_i, k_i, \lambda_i$ belonging to a family (A) with*

$$b_i = 4(r_i - \lambda_i), \qquad \ldots \text{(3.1)}$$

and N_i and N_i^* are the incidence matrices of D_i and D_i^*, where D_i^* is the complementary of $D, i = 1, 2$, then

$$N = N_1 \times N_2 + N_1^* \times N_2^* \qquad \ldots \text{(3.2)}$$

is the incidence matrix of a BIB design D with parameters v, b, r, k, λ given by

$$\left. \begin{array}{l} v = v_1 v_2, \ b = b_1 b_2, \\[1mm] r = r_1 r_2 + (b_1 - r_1)(b_2 - r_2) \\[1mm] k = k_1 k_2 + (v_1 - k_1)(v_2 - k_2) \\[1mm] \lambda = r - b/4 \end{array} \right\} \qquad \ldots \text{(3.3)}$$

and the design D also belongs to the same family (A).

Proof : For any $(0, 1)$ matrix A of order (m, n) define $\mathbf{A}$ to be the corresponding $(-1, 1)$ matrix obtained by changing 0 into -1. Then obviously

$$\mathbf{A} = 2A - J_{m,n} \qquad \ldots \text{(3.4)}$$

Hence
$$\mathbf{N_1} \times \mathbf{N_2} = (2N_1 - J_{v_1, b_1})\times(2N_2 - J_{v_2, b_2})$$
$$= 2[N_1 \times N_2 + (J_{v_1, b_1} - N_1)\times(J_{v_2, b_2} - N_2)] - J_{v_1 v_2, b_1 b_2}$$
$$= 2N - J_{v_1 v_2, b_1 b_2}. \qquad \ldots \text{(3.5)}$$

If we call $\mathbf{N_i}, i = 1, 2$ the $(-1, 1)$ incidence matrix of D_i, then it has been shown in by Sillitto (1957) that $\mathbf{N_1} \times \mathbf{N_2}$ is the $(-1, 1)$ incidence matrix of D if the conditions of Theorem 1 are satisfied. Hence (3.4), (3.5) imply that N is the incidence matrix of the design D. Obviously D belongs to the same family (A).

ON A TWO PARAMETER FAMILY OF BIBD

4. PARAMETERS OF THE FAMILY (A) OF BIB DESIGNS

In this section we derive the parameters of the family (A) in a form which is equivalent to that given by Sillitto (1957) but which is more directly useful to us. Let D belonging to the family (A) have parameters v, b, r, k, λ. In virtue of $i = 4(r-\lambda)$ and (2.1) we have

$$k = r(r-\lambda)/(r-2\lambda)^2, \quad v = \left[\frac{2(r-\lambda)}{(r-2\lambda)}\right]^2 \qquad \ldots \ (4.1)$$

Since v is an integer $2\lambda/(r-2\lambda)$ must be an integer which is either even or odd.

If $2\lambda/(r-2\lambda)$ is even $= 2m$, where m is an integer, then $r = \lambda(1+m)/m$. Since r is an integer we have $\lambda = mt$, where t is an integer. Then from (4.1) and (2.1) we get

$$v = [2(1+m)]^2, \ k = (1+m)(1+2m), \ r = t(1+2m), \ b = 4(1+m)t, \ \lambda = mt.$$

Putting $2(1+m) = s$, we get the sub-family of (A) with parameters

$$v = s^2, \ b = 2st, \ r = (s-1)t, \ k = s(s-1)/2, \ \lambda = \frac{(s-2)t}{2} \qquad \ldots \ (4.2)$$

where s and t are integers, s even and $2(s-1)t \geqslant s(s-1)$. Since b is positive s and t in (4.2) are either both positive or both negative. By simultaneously changing the signs of s and t in (4.2) we get

$$v = s^2, \ b = 2st, \ r = (s+1)t, \ k = s(s+1)/2, \ \lambda = (s+2)t/2 \qquad \ldots \ (4.3)$$

where $2(s+1)t \geqslant (s+1)$. The parameters (4.2) and (4.3) are parameters of complementary designs and given any one of them the other can always be constructed. Hence without loss of generality we may define the sub-family (A_1) of (A) by the integers s and t where the parameters of the design $A_1(s, t)$ are

$$v = s^2, \quad b = 2st, \quad r = (s-1)t, \quad k = s(s-1)/2, \quad \lambda = \frac{(s-2)t}{2} \qquad \ldots \ (4.4)$$

and s and t are positive integers, s even and $2t \geqslant s$.

Similarly if $2\lambda/(r-2\lambda)$ is odd $= 2m+1$ then $r = 4\lambda(1+m)/(1+2m)$. Since r is an integer, this implies that $\lambda = (1+2m)t$. Proceeding as before and putting $2m+3 = s$, where s is odd, we get the series with

$$v = s^2, \quad b = 4st, \quad r = 2(s-1)t, \quad k = s(s-1)/2, \quad \lambda = (s-2)t \qquad \ldots \ (4.5)$$

with $4(s-1)t \geqslant s(s-1)$. As before simultaneously changing the sign of s and t, we get the complementary design with

$$v = s^2, \quad b = 4st, \quad r = 2(s+1)t, \quad k = s(s+1)/2, \quad \lambda = (s+2)t \qquad \ldots \ (4.6)$$

with $4(s+1)t \geqslant s(s+1)$. Hence we can define the sub-series (A_2) of (A) given by the integers s and t where the parameters of $A_2(s, t)$ are given by

$$v = s^2, \quad b = 4st, \quad r = 2(s-1)t, \quad k = s(s-1)/2, \quad \lambda = (s-2)t, \qquad \ldots \ (4.7)$$

where s and t are positive integers, s odd and $4t \geqslant s$. Replacing s by $2s$ and t by s in (4.4) we get

$$v = b = 4s^2, \quad r = k = s(2s-1), \quad \lambda = s(s-1) \qquad \ldots \ (4.8)$$

which is the series given by Menon (1962).

5. Composition of members of the family (A)

Let $A_1(s_1, t_1)$ and $A_1(s_2, t_2)$ be two members of (A_1) given by parameters (4.4). Using the composition theorem it is easily seen that we get a BIB design whose complementary design is the design $A_1(s_1 s_2, 2t_1 t_2)$. The particular case of this result when $s_1 = 2t_1$ and $s_2 = 2t_2$ is given by Memon (1962) when the original designs are obtained by the method of difference sets.

In a similar manner it can be shown that the existence of $A_2(s_1, t_1)$ and $A_2((s_2, t_2)$ implies the existence of $A_3(s_1 s_2, 4t_1 t_2)$ and the existence of $A_1(s_1, t_1)$ and $A_2(s_2, t_2)$ implies the existence of $A_1(s_1 s_2, 4t_1 t_3)$.

Theorems 5 and 6 of Menon (1962) imply the existence of $A_1(2^m 6^n, 2^{m+n-1} 3^n)$ for non-negative integers m and n such that $m+n \geqslant 1$. The design with parameters (16, 24, 9, 6, 3) due to Bhattacharya (1944) is the design $A_1(4, 3)$ and the design corresponding to the finite Euclidean plane $EG(2, 3)$ is the design $A_2(3, 1)$. All these designs can be used to generate infinitely many members of the family (A). Further from the designs $A_1(4, 3)$ and $A_1(4, 2)$ by suitable repetitions we get $A_1(4, t)$ for all $t \geqslant 2$. An independent and nonisomorphic solution of $A_1(4, 5)$ is given by Rao (1961).

6. A family of symmetric BIB designs

If a symmetric BIB design with parameters (v, r, λ) exists, then by the process of block section (Bose, 1939) we get a BIB design with parameters $(v-r, v-1, r, r-\lambda, \lambda)$. If this belongs to the family (A), then $v-1 = 4(r-\lambda)$. Combined with $\lambda(v-1) = r(r-1)$, this shows that r is a perfect square and that $r = (r-2\lambda)^2$. Putting $r = s^2$, we have $\lambda = s(s\pm1)/2$, and hence we get two series with

$$v = s^2+(s+1)^2, \quad r = s^2, \quad \lambda = s(s-1)/2 \qquad \ldots \quad (6.1)$$

and

$$v = (s-1)^2+s^2, \quad r = s^2, \quad \lambda = s(s+1)/2. \qquad \ldots \quad (6.2)$$

By putting $s+1$ for s in (6.2) and taking the complementary design, we get a design with parameters (6.1).

From (6.1) and (6.2) using the methods of block section and block intersection (Bose, 1939) we get the following designs.

$$v = (s+1)^2, \quad b = 2s(s+1), \quad r = s^2, \quad k = s(s+1)/2, \quad \lambda = s(s-1)/2 \qquad \ldots \quad (6.3)$$

$$v = s^2, \quad b = 2s(s+1), \quad r = s^2-1, \quad k = s(s-1)/2, \quad \lambda = (s+1)(s-2)/2 \quad \ldots \quad (6.4)$$

$$v = (s-1)^2, \quad b = 2s(s-1), \quad r = s^2, \quad k = s(s-1)/2, \quad \lambda = s(s+1)/2 \qquad \ldots \quad (6.5)$$

$$v = s^2, \quad b = 2s(s-1), \quad r = s^2-1, \quad k = s(s+1)/2, \quad \lambda = (s+2)(s-1)/2. \ldots \quad (6.6)$$

All these four designs belong to the family (A). For example with s even (6.4) is $A_1(s, s+1)$, and (6.3) is $A_2(s+1, s/2)$ and the complementaries of (6.6) and (6.5) are $A_1(s, s-1)$ and $A_2(s-1, s/2)$. All these designs can, therefore, be constructed for any s if (6.1) and (6.2) exist for the corresponding values of s. For $s = 2$ and 3, (6.1) gives the symmetric designs (13, 4, 1) and (25, 9, 3) which are known to exist. It is easy to verify that if p is any odd prime factor of $r-\lambda$ in (6.1) or (6.2), then λ is either 0 or 1 (mod p). Hence the theorems of Chowla and Ryser (1950) on the impossibility

of symmetric BIB designs do not yield any positive result. For $s = 4, 5, 6, 7, 8, 9,$ 10 and 11, it can be proved that the corresponding residue difference sets (mod v) do not exist by using the results of Hall and Ryser (1951) and Hall (1956). It may be noted that the solutions to the designs (6.3)—(6.6) may exist even when the designs (6.1) and (6.2) are impossible.

7. Connection with Hadamard matrices

Let N denote the incidence matrix of the design (4.8) and let $\overline{N}$ denote the corresponding matrix by changing 0 into -1. Any two rows of N contain the ordered pairs (1, 1), (1, 0), (0, 1), (0, 0) exactly s^2-s, s^2, s^2 and s^2+s times respectively. In the corresponding two rows of $\overline{N}$ these pairs go into (1, 1) (1, -1), (-1, 1) and (-1, -1). Hence the scalar product of any two rows of $\overline{N}$ is obviously zero. Thus $\overline{N}$ is a Hadamard matrix of order $4s^2$. Hence we have the following.

Theorem 2 : The existence of a symmetric BIB design with parameters $(4s^2, 2s^2-s, s^2-s)$ implies the existence of a Hadamard matrix H_{4s^2}.

This theorem can also be derived in the following manner which incidentally provides a partial converse to the above theorem. Defining the distance between any two rows of a (0, 1) matrix as the number of places in which they differ it is easy to verify that

$$A = \begin{pmatrix} N \\ N^* \end{pmatrix}$$

represents a set of $8s^2$ rows with $4s^2$ components in 0 and 1 such that the distance between any two rows is $\geqslant 2s^2$. Applying Theorem 1 in Bose and Shrikhande (1959) we get the desired result. To actually construct the Hadamard matrix we proceed as follows. Replace every 0 by 1 in the first row of A, and in the corresponding columns of A interchange 0 and 1 to obtain the matrix B. Each column of A and hence of B contains each element 0 and 1 exactly $4s^2$ times. Let B_1 be the square submatrix of order $4s^2$ in B such that the first column of B_1 contains only the element 1. Replacing 0 by -1 in B_1, it is easy to verify as given by Bose and Shrikhande (1959) that we get a Hadamard matrix H_{4s^2}.

Now suppose H_{4s^2} exists. Without loss of generality we can take the first row and first column to consist entirely of unity. Then suppressing this row and column and changing -1 into 0 we get the incidence matrix N of a symmetric BIB design with parameters $(4s^2-1, 2s^2-1, s^2-1)$. Suppose this design satisfies the condition (C_s) that it is possible to find a set of $p = 2s^2-s-1$ columns and rows of N, such that in these p columns each one of the p rows has exactly s^2-s-1 positions 1 and each of the remaining $q = 2s^2+s$ rows has exactly s^2-1 positions occupied by 1. Without loss of generality, we can then put

$$N = \begin{pmatrix} N_1 & N_3 \\ N_2 & N_4 \end{pmatrix}$$

when N_1 and N_4 are square matrices of order p and q respectively. Further, each row of N_1, N_2, N_3 and N_4 contains 1 in exactly s^2-s-1, s^2-1, s^2+s and s^2 positions respectively. Now consider

$$M_1 = \begin{pmatrix} N_1 & N_3 \\ N_2^* & N_4^* \end{pmatrix}$$

It is easy to see that the distance δ between any two of the first p rows or any two of the last q rows of M_1 is $2s^2$, whereas the distance between two rows one from each of the two sets of p and q rows is $2s^2-1$. Since the distance between any two rows remains constant by interchanging 0 and 1 in any given columns, we see that the same result is true for

$$M_2 = \begin{pmatrix} N_1 & N_3^* \\ N_2^* & N_4 \end{pmatrix}$$

Obviously every row of $(N_1, \ N_3^*)$ contains 1 in $(s^2-s-1)+s^2 = 2s^2-s-1$ positions Similarly every row of $(N_2^* \ \ N_4)$ contains 1 in $(s^2-s)+s^2 = 2s^2-s$ places. Since the distance δ between two rows is $r_1+r_2-2\lambda$ where r_1 and r_2 are the number of 1's in the rows respectively and λ is the number of positions occupied by (1, 1), it is obvious that for the incidence matrix M_2, each of the first p treatments occurs $2s^2-s-1$ times and any two of the p treatments occur together $\dfrac{2(2s^2-s-1)-2s^2}{2} = s^2-s-1$ times. Similarly each of the q treatments occurs $2s^2-s$ times and any two of the q treatments occur together s^2-s times. Further any one of the p treatments occur with any one of the q treatments exactly s^2-s times. Noting that N_2^* contains 1 in s^2-s positions and regarding the rows and columns to correspond with blocks and treatments respectively, the matrix

$$M_3 = \begin{pmatrix} 1 & J_{1p} & 0_{1q} \\ J_{p1} & N_1 & N_3^* \\ 0_{q1} & N_2^* & N_4 \end{pmatrix}$$

can be regarded as giving an arrangement of $v = 4s^2$ treatments into v blocks each containing $k = 2s^2-s$ treatments such that any two blocks have exactly $\lambda = \dfrac{k(k-1)}{v-1} = s^2-s$ treatments in common. From Chowla and Ryser (1950) it follows that M_3 represents the incidence matrix of a symmetric BIB design with parameters (4.8). Changing s into $-s$, i.e., assuming the condition (C_{-s}) we similarly get a symmetric BIB design with parameters $(4s^2, 2s^2+s, s^2+s)$ and hence the complementary design given by parameters (4.8) Hence we have the following theorem.

Theorem 3 : *If there exists a symmetric BIB design with parameters* $(4s^2-1, 2s^2-1, s^2-1)$ *such that*

(i) *there exists a set of* $2s^2-s-1$ *blocks containing each of a set of* $2s^2-s-1$ *treatments* s^2-s-1 *times and each of the remaining* $2s^2+s$ *treatments* s^2-1 *times, or*

(ii) *there exists a set of $2s^2+s-1$ blocks containing each of a set of $2s^2+s-1$ treatments s^2+s-1 times and each of the remaining $2s^2-s$ treatments s^2-1 times, then we can construct a symmetric BIB design with parameters $(4s^2, 2s^2-s, s^2-s)$.*

Let e be an odd integer such that $e = p^m$, $e+2 = q^n$ where p and q are odd primes. Let x and y denote primitive elements of $GF(p^m)$ and $GF(q^n)$ respectively. Consider the set of elements (α, β) where $\alpha \epsilon GF(p^m)$ and $\beta \epsilon GF(q^n)$ the addition and multiplication being defined by

$$(\alpha_1, \beta_1)+(\alpha_2, \beta_2) = (\alpha_1+\alpha_2, \beta_1+\beta_2)$$

$$(\alpha_1, \beta_1)(\alpha_2, \beta_2) = (\alpha_1 \alpha_2, \beta_1 \beta_2).$$

Putting
$$z = (x, y), \quad w = (x, 0), \quad 0 = (0, 0)$$

we have the following theorem (Stanton and Sprott, 1958).

Theorem : (Stanton and Sprott). *The elements*

$$\left(z^0, z, z^2 \ldots, \ldots, z^{\frac{e^2-3}{2}}, 0, w^0, w^1, w^2 \ldots w^{e-2}\right)$$

form a difference set for a symmetric BIB design with parameters $(4s^2-1, 2s^2-1, s^2-1)$ *where* $2s = e+1$.

Since the first component of the elements $0, w^0, \ldots, w^{e-2}$ ranges over $GF(e)$, it is obvious that all the elements of the Galois Domain GD $e(e+2)$ are obtained by letting the second components in above elements range over $GF(e+2)$. Let $C_0, B_0, B_1, \ldots, B_{e-2}$ denote the set of $e+2 = 2s+1$ elements each obtained from $0, w^0, w^1, w^{e-2}$ respectively in this manner. The blocks and treatments of the design can then be divided into above subsets. Regarding the difference set as assigning the treatment $0 = (0, 0)$ to the blocks corresponding to the elements in the difference set, we can obtain the design by developing the initial row, corresponding to the treatment 0 by first developing according to the second component and then according to the first component. It is easy to see that the set of elements z^i, $i = 0, 1, \ldots \frac{e^2-3}{2}$ form a multiplication abelian group of order $\frac{e^2-1}{2}$ and that the elements of the difference set other than 0 contain exactly $\frac{e+3}{2} = s+1$ elements with the first component different from 0. Since the treatment 0 occurs only in the block $0 = (0, 0)$ of C_0, it is obvious that every treatment of C_0 occurs exactly once in the blocks of C_0. The treatment w^i occurs in those blocks obtained by adding x^i to the first component of each of the elements of the difference set. Since the difference set contains exactly $s+1$ elements with the first component $-x^i$, we see that the treatment w^i occurs in $s+1$ blocks of C_0 and hence each element of B_i also occurs in blocks of C_0 exactly $s+1$ times. Obviously $0 = (0, 0)$ is occurring in $s+1$ blocks of each B_i and hence each element of C_0 occurs in $s+1$ blocks of each B_i. Similarly each element of B_j, $j \neq i$ will occur in blocks B_i exactly $s+1$ times, whereas each treatment of B_i will

occur in these blocks exactly once. The number of times each treatment occurs in the blocks is, therefore, represented by the scheme

	C_0	B_0	B_1	.	.	.	B_{2s-3}
C_0	1	$s+1$	$s+1$	.	.	.	$s+1$
B_0	$s+1$	1	$s+1$	.	.	.	$s+1$
B_1	$s+1$	$s+1$	1	.	.	.	$s+1$
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
B_{2s-3}	$s+1$	$s+1$	$s+1$	.	.	.	1

where the rows and columns corresponds to sets of treatment and blocks respectively and each diagonal element is 1 whereas each nondiagonal element is $s+1$. Since each set of treatments and blocks represents a set of $2s+1$ entities, all that one has to do is to select a set of $s-1$ columns of the above matrix. The blocks thus obtained satisfy obviously condition (i) of Theorem 3. Hence we have the following :

Corollary : *If e and $e+2$ are both odd prime powers, then the design with parameters (4.8) exists with $s = \dfrac{e+1}{2}$.*

Taking $e = 9, 11, 17, 23, 25, 27, 29, 41, 47, 59, 71, 79$ we obtain that the design with parameters (4.8) exists for $s = 5, 6, 9, 12, 13, 14, 15, 22, 24, 30, 36, 40$. The case $s = 5$ corresponds to the design with parameters (100, 45, 20), which is mentioned by Hall (1956) as an unsolved case in difference sets. The results for $s = 6, 12, 24$ already follows from the results of Menon (1962). Theorem 1 can now be used to provide solutions for other values of s. For $s \leqslant 30$, it is then possible to obtain solutions for all values excepting $s = 7, 11, 17, 22, 23, 25, 27$ and 29.

REFERENCES

BHATTACHARYA, K. N. (1944) : On a new symmetrical balanced incomplete block design. *Bull. Cal. Math. Soc.*, **26**, 91-96.

BOSE, R. C. (1939) : On the construction of balanced incomplete block designs. *Ann Eugen.*, **9**, 353-399.

BOSE, R. C. and SHRIKHANDE, S. S. (1959) : A note on a result in the theory of code construction. *Information and Control*, **2**, 183-194.

CHOWLA, S. and RYSER, H. J. (1950) : Combinatorial problems. *Can. J. Math.* **2**, 95-99.

FISHER, R. A. (1940) An examination of different possible solutions of a problem in incomplete blocks. *Ann. Eugen.*, **10**. 52-75.

HALL, MARSHALL Jr. (1956) : A survey of difference sets. *Proc. Amer. Math. Soc.*, **7**, 975-986.

HALL, MARSHALL Jr. and RYSER, H. J. (1951) : Cyclic incidence matrices., *Can. J. Math.*, **3**, 495-502.

MENON, P. KESHAV (1962) : On difference sets whose parameters satisfy a certain relation. To appear in *Proc. Amer. Math. Soc.*

RAO, C. R. (1961) : A study of BIB designs with replications 11 to 15. *Sankhyā*, Series A, **23**, 117-127.

SILLITTO, G. P. (1957) : An extension property of a class of balanced incomplete block designs. *Biometrika*, **44**, 278-279.

STANTON, R. G. and SPROTT, D. A. (1958) : A family of difference sets. *Can. J. Math.*, **10** 73-77.

VARTAK, M. N. (1955) : On an application of Kronecker product of matrices to statistical designs. *Ann. Math. Stat.*, **26** 420-438.

Paper received : October, 1961.

DECOMPOSITION BY BILATERAL COSETS AND ITS GENERALIZATION

By MOTOSABURO MASUYAMA

Tokyo University

and

Indian Statistical Institute, Calcutta

SUMMARY. In previous papers the author (Masuyama, 1961a, 1961b, 1961c and 1961d) introduced (1) the factorial decomposition, (2) the hierarchic decomposition, (3) the symmetric decomposition and (4) the periodic decomposition of a finite module or a finite ring with unity. Each decomposition supplies a family of Partially Balanced Incomplete Block designs, if we make one element of a module or of a ring correspond to one variety and vice versa. As is proved by the author (Masuyama, 1961d) the periodic decomposition is a refinement of other decomposition. We shall generalize in Section 1 the concept of the periodic block in the sense that ordinary cosets are special cases of bilateral or double cosets in the theory of groups and shall generalize, further, from the viewpoint of mapping in Section 2 which was noticed by P. K. Menon.*

1. Suppose, $\mathscr{A}$ be a ring of order v with unity e. Let $\mathscr{G}$ and $\mathscr{H}$ be two multiplicative groups of order g and h respectively contained in this ring, e being the unity of $\mathscr{G}$ and $\mathscr{H}$. $\mathscr{G}$ and $\mathscr{H}$ are not necessarily distinct. a_i, G_j and H_l being an element of $\mathscr{A}$, $\mathscr{G}$ and $\mathscr{H}$ respectively, $G_j a_i H_l$ is an element of $\mathscr{A}$. We arrange this element in the row (a_i) and in the column (G_j, H_l), for $i = 1, 2, ..., v$, $j = 1, 2, ..., g$, $l = 1, 2, ..., h$. The order of arranging the heading (a_i) or (G_j, H_l) is immaterial, so far as all possible cases occur just once.

TABLE 1

	$(G_1, H_1) = (e, e)$	$\therefore$	(G_j, H_l)	...	(G_g, H_h)
$(a_1) = (e)$	eee	$\therefore$	$G_j e H_l$	...	$G_g e H_h$
$\vdots$	$\vdots$		$\vdots$		$\vdots$
(a_i)	$e a_i e$	...	$G_j a_i H_l$	...	$G_g a_i H_h$
$\vdots$	$\vdots$		$\vdots$		$\vdots$
(a_v)	$e a_v e$	...	$G_j a_v H_l$	...	$G_g a_v H_h$

Then each column contains all elements of the ring exactly once, because if we have

$$G_j a H_l = G_j b H_l, \qquad \ldots \text{(1)}$$

then multiplying G_j^{-1} from the left and H_l^{-1} from the right we get

$$a = b. \qquad \ldots \text{(2)}$$

Identical elements of the ring may appear in the same row more than once. Suppose that there are exactly d elements contained in the row (a) which are equal to a, i.e.

$$a = G_{j_2} a H_{l_2} = G_{j_3} a H_{l_3} = \cdots = G_{j_d} a H_{l_d}. \qquad \ldots \text{(3)}$$

(i) If $d = gh$, we have $\qquad a = G_j a H_l \qquad \ldots \text{(4)}$

for $\quad j = 1, 2, \ldots, g; \quad l = 1, 2, \ldots, h.$

(ii) If $d < gh$, there are elements which are not equal to a. Let any one of them be GaH. Then multiplying G from the left and H from the right we have by (3)

$$GaH = GG_{j_2} a H_{l_2} H = \cdots = GG_{j_d} a H_{l_d} H, \qquad \ldots \text{(5)}$$

with $(G, H) \neq (GG_{j_m}, H_{l_m} H)$ for $m = 2, 3, \ldots, d$. Thus GaH reduced to an element of $\mathscr{A}$ appears at least d times on the same row (a). If there were more than d elements which are equal to GaH, let $G_0 a H_0$ be any one of them.

Then we would have

$$a = G_{j_2} a H_{l_2} = \cdots = G_{j_d} a H_{l_d} = G^{-1} G_0 a H_0 H^{-1} \qquad \ldots \text{(6)}$$

with $(G_{j_m}, H_{l_m}) \neq (G^{-1} G_0, H_0 H^{-1})$ for $m = 2, 3, \ldots, d$, which is contrary to our assumption. Thus each distinct element of $\mathscr{A}$ is contained exactly d times in the row (a), if it is contained in the row (a).*

A block which contains all elements of one row of Table 1 is called a bilateral coset block or in short BC block and a block which contains every different element in the same row exactly once is called a normalized BC block. Then two BC blocks obtained from different rows are either identical or disjoint.

If $\qquad G_j a H_l = G_m b H_n, \quad \text{for} \quad a \neq b, \qquad \ldots \text{(7)}$

we have $\quad GaH = GG_j^{-1} G_m b H_n H_l^{-1} H, \quad \text{for any } G \epsilon \mathscr{G} \text{ and } H \epsilon \mathscr{H}. \qquad \ldots \text{(8)}$

Thus all elements GaH are contained in the row (b). Similarly, all elements GbH are contained in the row (a). Therefore, the sum of all possible non-identical normalized BC blocks contains all elements of the ring once and only once.

---

* The author wishes to express his thanks to Dr. P. K. Menon for kindly pointing out the mistake in the proof on this point in the first manuscript.

We shall now prove that a product of any two BC blocks is represented by a linear combination of BC blocks, coefficients being non-negative integers. In fact we have

$$G_j a H_l + G_m b H_n = G_j(a + G_j^{-1} G_m b H_n H_l^{-1}) H_l$$
$$= G_j(a + G_p b H_q) H_l, \qquad \qquad \dots \ (9)$$

in which the second term in the bracket runs through every element of the row (b) once and only once, for all combinations of m and n, whatever j and l may be so far as these two suffixes are fixed.

$$c_{pq} = a + C_p b H_q \qquad \qquad \dots \ (10)$$

being an element of the ring $\mathcal{A}$, the block

$$\{G_1 c_{pq} H_1, \ \dots, \ G_j c_{pq} H_l, \ \dots, \ G_g c_{pq} H_h\} \qquad \qquad \dots \ (11)$$

for fixed values of p and q is a BC block.   q.e.d.

The periodic block is a special case of our bilateral coset block in which one of $\mathcal{G}$ and $\mathcal{H}$ consists of only one element e.

2.   The above result is easily generalized, if we realize that the essential features of the above proof are (i) the multiplicative group property of the transformations or mappings τ_{jl} of an element of a, i.e. $G_j a H_l$ in this case, and (ii) the distributive property of the transformation or the isomorphism between a and $\tau_{jl} a$. The first property is used in getting the formulas, (2), (5), (6), and (8) and the second one is used in getting the formula (9). The existence of the unity in $\mathcal{A}$, which we have assumed in Section 1, is needed only when we utilize the group property of the mappings.

Now let $\mathcal{M}$ be a finite module of order v and $\tau_j a$ be a mapping of a in $\mathcal{M}$ into $\mathcal{M}$. Then all mappings τ_j constitute a semi-group $\mathcal{T}(\mathcal{M})$, of which an inversible element gives a bijective mapping. All the bijective mappings constitute a symmetric group, i.e. a substitution group $\mathfrak{S}(\mathcal{M})$, of which automorphic mappings constitute a subgroup $\mathfrak{A}(\mathcal{M})$ of $\mathfrak{S}(\mathcal{M})$. Any subgroup of $\mathfrak{A}(\mathcal{M})$ can be used for generating a Partially Balanced Incomplete Block design. Table 1 in Section 1 is replaced by the following table :

<div align="center">TABLE 2</div>

	τ_1	$\cdots$	τ_j	$\cdots$	τ_g
(a_1)	$\tau_1 a_1$	$\cdots$	$\tau_j a_1$	$\cdots$	$\tau_g a_1$
$\vdots$	$\vdots$		$\vdots$		$\vdots$
(a_i)	$\tau_1 a_i$	$\cdots$	$\tau_j a_i$	$\cdots$	$\tau_g a_i$
$\vdots$	$\vdots$		$\vdots$		$\vdots$
(a_v)	$\tau_1 a_v$	$\cdots$	$\tau_j a_v$	$\cdots$	$\tau_g a_v$

The set of all elements in a row constitute a block, in which all distinct elements appear with the same frequency, say d-times. A block which is derived from one row and contains all distinct elements exactly once, is called a normalized block. The block thus obtained may be qualified by the specific mapping used.

3. We shall illustrate our method by one of the commonest transformations, i.e. by conjugation. Suppose that a is an element of a ring $\mathscr{A}$ of order v with unity and G_j is an element of a multiplicative group of order g contained in $\mathscr{A}$. Then the conjugated block

$$\{G_1 a G_1^{-1}, \ldots, G_j a G_j^{-1}, \ldots, G_g a G_g^{-1}\} \qquad \ldots \ (12)$$

or its sum generates a Partially Balanced Incomplete Block design by multiplying $\{s\}$, s being every element of $\mathscr{A}$. The mapping 'conjugation' satisfies two conditions stated in Section 2.

Example : Consider the matrix ring of order 16 (see Appendix) which is quoted by (Masuyama, 1961d). Let $\mathscr{G}$ be (12, 23, 31, 32, 21, 13), of which (12, 23, 31), (12, 13), (12, 21) and (12) are its subgroups with regard to multiplication. The normalized conjugated blocks obtained by $\mathscr{G}$ are as follows :

$$E = \{00\},$$
$$A = \{12\},$$
$$B = \{23, 31\},$$
$$C = \{32, 21, 13\},$$
$$D = \{01, 20, 33\},$$
$$F = \{02, 03, 10, 11, 30, 22\}.$$

All these blocks are self-conjugate. The multiplication table of these blocks are given by Table 3.

TABLE 3

	E*	A*	B*	C*	D*	F*
E	E					
A	A	E				
B	B	B	2E+2A			
C	C	D	F	3E+2C		
D	D	C	F	3A+2D	3E+2C	
F	F	F	2C+2D	3B+2F	3B+2F	6E+6A+4C+4D

There are simple linear relations among these blocks and the periodic blocks (Masuyama, 1961d), i.e.

$$F_1 = A+B$$
$$F_2+F_3+F_4 = D+F$$

and
$$F_5 = C.$$

By setting
$$D_1 = A+C+F \quad \text{and} \quad D_2 = B+D$$

we obtain the following multiplication table :

TABLE 4

	E^*	D_1^*	D_2^*
E	E		
D_1	D_1	$10E+6(D_1+D_2)$	
D_2	D_2	$3D_1+4D_2$	$5E+2D_1$

The initial block D_1 and the initial block $(E+D_2)$ supply two Balanced Incomplete Block designs which are complementary. The initial block D_2 supplies a Partially Balanced Incomplete Block design with the following parameters of the first kind :

$$v = b = 16, \quad k = r = 5, \; n_1 = 10, \; n_2 = 5, \lambda_1 = 2 \text{ and } \lambda_2 = 0.$$

The parameters of the second kind are given in Table 4. A Partially Balanced Incomplete Block design with parameters of the same first and second kinds is given by Clatworthy (1956). However, his design is not cyclic.

Appendix : Matrix ring of order 16

THE ADDITION TABLE

	00	01	02	03	10	11	12	13	20	21	22	23	30	31	32	33
00	00	01	02	03	10	11	12	13	20	21	22	23	30	31	32	33
01	01	00														
02	02	03	00													
03	03	02	01	00												
10	10	11	12	13	00											
11	11	10	13	12	01	00										
12	12	13	10	11	02	03	00									
13	13	12	11	10	03	02	01	00								
20	20	21	22	23	30	31	32	33	00							
21	21	20	23	22	31	30	33	32	01	00						
22	22	23	20	21	32	33	30	31	02	03	00					
23	23	22	21	20	33	32	31	30	03	02	01	00				
30	30	31	32	33	20	21	22	23	10	11	12	13	00			
31	31	30	33	32	21	20	23	22	11	10	13	12	01	00		
32	32	33	30	31	22	23	20	21	12	13	10	11	02	03	00	
33	33	32	31	30	23	22	21	20	13	12	11	10	03	02	01	00

THE MULTIPLICATION TABLE

L / R	00	01	02	03	10	11	12	13	20	21	22	23	30	31	32	33
00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00
01	00	00	00	00	01	01	01	01	02	02	02	02	03	03	03	03
02	00	01	02	03	00	01	02	03	00	01	02	03	00	01	02	03
03	00	01	02	03	01	00	03	02	02	03	00	01	03	02	01	00
10	00	00	00	00	10	10	10	10	20	20	20	20	30	30	30	30
11	00	00	00	00	11	11	11	11	22	22	22	22	33	33	33	33
12	00	01	02	03	10	11	12	13	20	21	22	23	30	31	32	33
13	00	01	02	03	11	10	13	12	22	23	20	21	33	32	31	30
20	00	10	20	30	00	10	20	30	00	10	20	30	00	10	20	30
21	00	10	20	30	01	11	21	31	02	12	22	32	03	13	23	33
22	00	11	22	33	00	11	22	33	00	11	22	33	00	11	22	33
23	00	11	22	33	01	10	23	32	02	13	20	31	03	12	21	30
30	00	10	20	30	10	00	30	20	20	30	00	10	30	20	10	00
31	00	10	20	30	11	01	31	21	22	32	02	12	33	23	13	03
32	00	11	22	33	10	01	32	23	20	31	02	13	30	21	12	03
33	00	11	22	33	11	00	33	22	22	33	00	11	33	22	11	00

(i, j) stands for $\begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}$ with $i = c_{11} + 2c_{12}$ and $j = c_{21} + 2c_{22}$, and c_{kl} is an element of the module to modulo 2, i. e. 0 or 1.

References

Bourbaki, N. (1936) : *Eléments de Mathématique*, Hermann, Paris.

Clatworthy, W. H. (1956) : Contributions on partially balanced incomplete block designs with two associate classes. N.B.S., Applied Mathematics Series, No. 47.

Dubreil, P. (1954) : *Algébre*, Tome 1, Gauthier-Villars, Paris.

Hall, M. (1959) : *The Theory of Groups*, MacMillan.

Iyanaga, S. and Kodaira, K. (1961) : *Survey of Modern Mathematics*, I, Iwanami., Tokyo.

Lyapin, E. S. (1960) : *Polugruppy*, Gos. Izd. Fiz.—Mat. Lit., Moskva.

Masuyama, M. (1961a) : Calculus of blocks and a class of partially balanced incomplete block designs. *Rep. Stat. Appl. Res.*, **8**, 56-69.

——— (1961b) : On cyclic difference sets which generate orthogonal arrays, Part II. *Rep. Stat. Appl. Res.*, **8**, 70-76.

——— (1961c) : Le calcul des blocs et ses applications aux plans d'expérience, Colloques Internationaux de la Recherche Scientifique, Le Plan d'Expérience, 1-9.

——— (1961d) : La décomposition périodique dans le calcul des blocs—le raffinement des décompositions, *Bulletin de l'Institut International de Statistique*, **33e**, Session, Paris, 1-8.

Paper received : October, 1961.

CUMULANTS OF FUNCTIONS OF RANDOM VARIABLES

By G. S. JAMES and ALAN J. MAYNE

University of Leeds

SUMMARY. A combinatorial method is described of obtaining the joint cumulants of an arbitrary number of functions of an arbitrary number of variables, the cumulants of the variables being known. The important case when all r-th order cumulants of the independent variables are of the $(r-1)$-th order of small quantities is treated in detail, and explicit results are given as far as the fourth order of small quantities (at least). Proofs are only outlined, since they are implicit in previous work.

Let x be a random variable all of whose cumulants κ_r exist. Let y be a function of x expansible in a Taylor series,

$$y = A_0 + A_1 x + A_2 x^2 + \cdots, \qquad \qquad \ldots \quad (1)$$

Then, under certain circumstances, and in particular when y is a polynomial in x, all the cumulants K_r of y will exist. When y is a polynomial, they may be evaluated, via the moments, using (1). For example, if $A_3 = A_4 = \cdots = 0$ we have

$$K_1 = A_0 + A_1 \kappa_1 + A_2 (\kappa_2 + \kappa_1^2),$$
$$K_2 = A_1^2 \kappa_2 + 2 A_2 A_1 (\kappa_3 + 2\kappa_2 \kappa_1) + A_2^2 (\kappa_4 + 4\kappa_3 \kappa_1 + 2\kappa_2^2 + 4\kappa_2 \kappa_1^2),$$

and so on. When y is not a polynomial the cumulants K_r may not exist,[*] but even if they do their expressions will be complicated infinite series involving all the A_r and κ_r. To get usable results we assume that each κ_r is of order ν^{-r+1} in some large number ν, that the A_r are independent of ν (or at least of order ν^0), and in the expressions for the K_r we retain terms only up to a given order in ν.

In a previous article (James, 1955) one of us outlined a combinatorial method which can be used to carry out the calculations, and used it to prove the non-trivial result that K_r is also of order ν^{-r+1}. The proof depended, in what may seem rather a weird way, on a result published three years later (James, 1958; referred to in the previous paper as "James, 1955, not yet published"). Here we explain the method in more detail, sketch a complete proof, and provide explicit results as far as terms of order ν^{-4} (together with the terms of order ν^{-5} in K_6).

It transpires that it is simpler to generalize (1), by considering q variables $y^1, \ldots, y^q$, expressed in terms of p variables $x^1, \ldots, x^p$, by series of the type

$$y^a = A^a + A_i^a x^i + A_{ij}^a x^i x^j + \ldots \quad (A_{ij}^a = A_{ji}^a, \text{ etc.}). \qquad \ldots \quad (2)$$

The superfixes are indices, not exponents, and we use throughout the convention that terms with repeated indices, such as $i, j, \ldots$ above, are to be summed over $i, j, \ldots = 1, \ldots, p$. We shall use a slightly modified form of the "tensor" notation suggested

---

[*] *Sufficient* conditions for their existence are that x is a bounded random variable, whose bounds are numerically less than the radius of convergence of (1). It is *not* sufficient that y is an integral (entire) function of x, as is seen by taking x to be a standard normal deviate and $y = \exp x^2$.

by Kaplan (1952), in which the first, second, ... order multivariate cumulants of the x^i are κ^i, κ^{ij}, ... $(i, j, ... = 1, ..., p)$, and those of the y^a are K^a, K^{ab}, ... $(a, b, ... = 1, ..., q)$. For example

$$\kappa^i = \mathcal{E} \, x^i$$

and $$\kappa^{ij} = \mathcal{E} \, x^i x^j - (\mathcal{E} \, x^i)(\mathcal{E} \, x^j) = \text{cov} \, (x^i, \, x^j) \qquad [= \text{var} \, x^i \text{ if } i = j]. \quad ... \quad (3)$$

It is assumed that all the r-th order cumulants are of order v^{-r+1}. Results for the univariate case (1) can be obtained by taking $p = q = 1$ and writing κ_1, κ_2, ... for κ^1, κ^{11}, ..., and similarly for the K's.

We now write, temporarily, z_0, z_1, ... for the expressions 1, x^1, ..., x^p, $x^1 x^1$, $x^1 x^2$, ..., $x^p x^p$, $x^1 x^1 x^1$, ..., and rewrite (2) in the form

$$y^a = A^{a\lambda} \, z_\lambda. \qquad ... \quad (4)$$

It follows immediately from the relations $\mathcal{E} y^a = A^{a\lambda} \, \mathcal{E} z_\lambda$, $\mathcal{E} y^a y^b = A^{a\lambda} A^{b\mu} \, \mathcal{E} z_\lambda z_\mu$, ..., and from moment-cumulant relations analogous to (3), that

$$K^a = A^{a\lambda} \, \kappa_\lambda(z), \qquad K^{ab} = A^{a\lambda} A^{b\mu} \, \kappa_{\lambda\mu}(z), \, ... \qquad ... \quad (5)$$

Replacing the z's by their meanings as x-products, and introducing a notation in which, for example, $\kappa[ij, k]$ denotes the second-order cumulant of $x^i x^j$ and x^k, we have, for example

$$K^{abc} = A^a_i A^b_j A^c_k \kappa \, [i, j, k] + \overset{3}{\Sigma} A^a_{ij} A^b_k \, A^c_l \, \kappa \, [ij, k, l] + (\overset{3}{\Sigma} A^a_{ijk} A^b_l \, A^c_m \, \kappa \, [ijk, l, m]$$

$$+ \overset{3}{\Sigma} A^a_{ij} A^b_{kl} A^c_m \kappa \, [ij, kl, m]) + (\overset{3}{\Sigma} A^a_{ijkl} A^b_m A^c_n \kappa \, [ijkl, m, n]$$

$$+ \overset{6}{\Sigma} A^a_{ijk} A^b_{lm} A^c_n \kappa \, [ijk, lm, n] + A^a_{ij} A^b_{kl} A^c_{mn} \, \kappa \, [ij, kl, mn]) + ... \qquad ... \quad (6)$$

In this expression we have written

$$\overset{3}{\Sigma} A^a_{ij} A^b_k A^c_l \kappa \, [ij, k, l] \text{ for } (A^a_{ij} A^b_k A^c_l + A^b_{ij} A^a_k A^c_l + A^c_{ij} A^a_k A^b_l) \, \kappa \, [ij, k, l],$$

the sum over all (three) distinct terms obtained by permuting a, b, c. Note that

$$A^a_{ij} A^c_k A^b_l \kappa \, [ij, k, l] = A^a_{ij} A^b_k A^c_l \, \kappa \, [ij, k, l];$$

thus the interchange of superfixes of A coefficients possessing equal numbers of suffixes produces no new terms. Hence, if a particular summation is of terms which have ρ_1 A's with r_1 suffixes, ..., ρ_R A's with r_R suffixes $(r_1 > r_2 > ... > r_R)$ then there are

$$\rho! / \rho_1! ... \rho_R! \qquad ... \quad (7)$$

terms in the sum, where $\rho = \Sigma \rho_k$.

It remains to express the cumulants $\kappa[\ ]$ of x-products in terms of those of the x^i. We proceed via the corresponding moments $\alpha[\ ]$. We have, for example,

$$\alpha[ij, k, l] = \mathcal{E}(x^i x^j) x^k x^l = \mathcal{E} x^i x^j x^k x^l$$

$$= \kappa^{ijkl} + \overset{4}{\Sigma} \kappa^{ijk} \kappa^l + \overset{3}{\Sigma} \kappa^{ij} \kappa^{kl} + \overset{6}{\Sigma} \kappa^{ij} \kappa^k \kappa^l + \kappa^i \kappa^j \kappa^k \kappa^l \, ; \qquad ... \quad (8)$$

$$\kappa[ij, k, l] = \alpha[ij, k, l] - 1! (\alpha[ij, k] \alpha[l] + \alpha[ij, l] \alpha[k] + \alpha[k, l] \, \alpha[ij]) + 2! \alpha[ij] \, \alpha[k] \alpha[l] \quad ... \quad (9)$$

$$= \kappa^{ijkl} + \overset{2}{\Sigma} \kappa^{ikl} \kappa^j + \overset{2}{\Sigma} \kappa^{ik} \kappa^{jl}. \qquad ... \quad (10)$$

We can summarize the terms in (8) by means of the bipartitions, or arrays,

$$
\begin{array}{c|c}
ij & ij \\
k & k \\
l & l \\
\hline
1 & ijkl
\end{array}
\quad
\begin{array}{c|c}
ij & ij \\
k & k \\
l & \\
& l \\
\hline
2 & ijk \quad l
\end{array}
\quad
\begin{array}{c|c|c}
ij & i & j \\
k & k & \\
l & l & l \\
\hline
2 & ikl & j
\end{array}
\quad
\begin{array}{c|c}
ij & ij \\
k & \\
l & \\
& k \\
& l \\
\hline
1 & ij \quad kl
\end{array}
\quad
\begin{array}{c|c|c}
ij & i & j \\
k & k & \\
l & & l \\
\hline
2 & ik & jl
\end{array}
$$

$$
\text{I} \qquad \text{IIa} \qquad \text{IIb} \qquad \text{IIIa} \qquad \text{IIIb}
$$

$$
\begin{array}{c|c}
ij & ij \\
k & \\
l & \\
& k \\
& l \\
\hline
1 & ij \quad k \quad l
\end{array}
\quad
\begin{array}{c|c|c}
ij & i & j \\
k & k & \\
l & & l \\
\hline
4 & ik & j \quad l
\end{array}
\quad
\begin{array}{c|c}
ij & ij \\
k & k \\
l & l \\
\hline
1 & kl \quad ij
\end{array}
\quad
\begin{array}{c|c|c}
ij & i & j \\
k & & k \\
l & & l \\
\hline
1 & i & j \quad k \quad l
\end{array}
$$

$$
\text{IVa} \qquad\qquad \text{IVb} \qquad\qquad \text{IVc} \qquad\qquad \text{V}
$$

in which row "totals" correspond to the left-hand side, column totals to the various terms on the right, and the italicized numerals show the number of terms of each type (Fisher, 1930; Kaplan, 1952*). It should be noted that the numerical coefficients in (8) are all unity, and that there is one term corresponding to every possible way of splitting the row totals. When we subtract the three terms in parenthesis in (9), this removes the terms in (8) corresponding to arrays IIa, IIIa, IVb, IVc, which are *dissectable* or *unconnected*, and fall into two blocks. Terms corresponding to IVa and V, which fall into three blocks, are subtracted three times, but this is put right when we add $2\alpha[ij]\,\alpha[k]\,\alpha[l] = 2(\kappa^{ij}+\kappa^i\kappa^j)\kappa^k\kappa^l$. The net result is that in (10) we retain only terms corresponding to *connected* or *non-dissectable* arrays.

That this cancellation will always take place exactly is not quite obvious. The proof is mainly a matter of devising a suitable notation, and is given in James (1958) (Theorem 6.1). Essentially one uses relations of the type

$$\alpha[ij,\,k,\,l] = \kappa[ij,\,k,\,l] + (\kappa[ij,\,k]\,\kappa[l] + \kappa[ij,\,l]\,\kappa[k] + \kappa[k,\,l]\,\kappa[ij]) + \kappa[ij]\kappa[k]\kappa[l], \quad \dots \quad (11)$$

in which all coefficients are $+1$, and proceeds by induction on the order of the highest cumulant involved.

In order to obtain the K's (to a given order in ν) in finite terms, we shall assume that

$$\kappa^i = 0 \qquad (i = 1, \dots, p). \qquad\qquad \dots \quad (12)$$

More generally, we could take, say, $\kappa^i = 0(\nu^{-1})$, but the results would be more complicated. With this proviso, we see that to determine all terms of order ν^{-k} in a cumu-

---

*We have, however, reversed the row-column convention, as used by these authors. The present convention is more in accordance with matrix notation.

lant $K^{a \cdots b}$ of order ρ, we have to consider all arrays which satisfy the following conditions.

(i) Number of rows $= \rho$.

(ii) (Number of letters)$-$(number of columns) $= \Sigma$ (number of letters in column totals-1) $= k$.

(iii) There are at least two letters in each column total.

(iv) Array is connected.

Of course arrays obtained by mere permutations between rows or between columns do not count afresh. For example, the arrays giving terms of order ν^{-2} in K^{abc} are

$$
\begin{array}{cc|cc}
i & i & ik & i \quad k \\
j & j & j & j \\
k & k & l & \quad l \\
\hline
1 \; \, ijk & & 2 \; \, ij \; \; kl
\end{array}
$$

and

$$
K^{abc} = \{A_i^a A_j^b A_k^c \kappa^{ijk} + 2 \overset{3}{\Sigma} A_{ik}^a A_j^b A_l^c \kappa^{ij}\kappa^{kl}\} + O(\nu^{-3}).
$$

The factor 2 in the second term arises because

$$
A_{ik}^a A_j^b A_l^c \overset{2}{\Sigma} \kappa^{ij}\kappa^{kl} = A_{ik}^a A_j^b A_l^c (\kappa^{ij}\kappa^{kl} + \kappa^{il}\kappa^{jk}) = 2 A_{ik}^a A_j^b A_l^c \kappa^{ij}\kappa^{kl}.
$$

Similar, though not always complete, conversions of the inner summation into a numerical factor occur in other terms. The general apportionment between N and S, the numerical factor and the number of terms in the sum, may be set out as follows.

We call the result of replacing each element of an array by the number of letters it contains its *univariate image*. For example,

$$
\begin{array}{ccc|ccc}
ijl & ij & l & 3 & 2 & 1 \\
kmn & k & m \quad n \to & 3 & 1 & 1 \; \bar{1} \; , \\
r & & r & 1 & & 1 \\
\hline
& ijk \;\; lm \;\; nr & & & 3 \; 2 \; 2
\end{array}
\qquad
\begin{array}{ccc|ccc}
ijkm & ij & k & m & 4 & 2 \; 1 \; 1 \\
lnrs & l & n & rs \to & 4 & 1 \; 1 \; 2 \\
\hline
& ij \;\; kl \;\; mn \;\; rs & & & 2 \; 2 \; 2 \; 2
\end{array}
$$

In this paragraph the expressions "array", "row", "column" and "element" shall always refer to the image array. Two rows (or columns) are said to be *identical* if they contain the same elements in the same positions. Two arrays are *doubly related* if one can be obtained from the other both by a permutation of rows only and by a permutation of columns only. (It follows that their total rows and total columns are identical.) The *degeneracy* of an array, denoted by D, is the number of distinct arrays (including itself) which are doubly related to it. The numerical factor N is the product of ρ multinomial coefficients, one corresponding to each row, divided by a product

of factorials, each corresponding to a group of identical columns. The number of terms in the sum, S, is $\rho!$ divided by a product of factorials, each corresponding to a group of identical rows, and also by D. For the two arrays quoted we have respectively $D = 1$, $N = 3.6.1/1!1!1! = 18$, $S = 3!/1!1!1!D = 6$ and $D = 2$, $N = 12.12/1!2!1! = 72$, $S = 2!/1!1!D = 1$.

A connected array is, by definition, one in which it is possible to proceed from any element to any other, along rows and columns, without passing over an empty place. Consequently, it may be built up, letter by letter, in such a way that each new letter (except the first) starts a new row, a new column, or neither, but never both. Hence, if there are ρ rows, τ columns and w letters in the array, we have

$$w \geqslant 1+(\rho-1)+(\tau-1) = \rho+\tau-1. \qquad \ldots \quad (13)$$

Thus the order of magnitude of the κ product, which is precisely $\nu^{-k} = \nu^{-w+\tau}$, is $\nu^{-\rho+1}$ or lower, and so the same is true for $K^{a \cdots b}$.

Results

We enclose terms of orders ν^0, ν^{-1}, ... in braces, thus : $_0\{\ \}$, $_1\{\ \}$, ...

$$K^a = {}_0\{A^a\} + {}_1\{A^a_{ij}\kappa^{ij}\} + {}_2\{A^a_{ijk}\kappa^{ijk} + 3A^a_{ijkl}\kappa^{ij}\kappa^{kl}\}$$

$$+ {}_3\{A^a_{ijkl}\kappa^{ijkl} + 10A^a_{ijklm}\kappa^{ijk}\kappa^{lm} + 15A^a_{ijklmn}\kappa^{ij}\kappa^{kl}\kappa^{mn}\}$$

$$+ {}_4\{A^a_{ijklm}\kappa^{ijklm} + 15A^a_{ijklmn}\kappa^{ijkl}\kappa^{mn} + 10A^a_{ijklmn}\kappa^{ijk}\kappa^{lmn}$$

$$+ 105A^a_{ijklmnr}\kappa^{ijk}\kappa^{lm}\kappa^{nr} + 105A^a_{ijklmnrs}\kappa^{ij}\kappa^{kl}\kappa^{mn}\kappa^{rs}\} + O(\nu^{-5}),$$

$$K^{ab} = {}_1\{A^a_i A^b_j \kappa^{ij}\} + {}_2\{\overset{2}{\Sigma} A^a_{ij}A^b_k \kappa^{ijk} + (3\overset{2}{\Sigma} A^a_{ijk}A^b_l + 2A^a_{ik}A^b_{jl})\kappa^{ij}\kappa^{kl}\}$$

$$+ {}_3\{(\overset{2}{\Sigma} A^a_{ijk}A^b_l + A^a_{ij}A^b_{kl})\kappa^{ijkl} + (4\overset{2}{\Sigma} A^a_{ijkl}A^b_m + 6\overset{2}{\Sigma} A^a_{ijlm}A^b_k$$

$$+ 6\overset{2}{\Sigma} A^a_{ijl}A^b_{km} + 3\overset{2}{\Sigma} A^a_{ilm}A^b_{jk})\kappa^{ijk}\kappa^{lm} + (15\overset{2}{\Sigma} A^a_{ijklm}A^b_n + 12\overset{2}{\Sigma} A^a_{ijkm}A^b_{ln}$$

$$+ 9A^a_{ijk}A^b_{lmn} + 6A^a_{ikm}A^b_{jln})\kappa^{ij}\kappa^{kl}\kappa^{mn}\} + {}_4\{(\overset{2}{\Sigma} A^a_{ijkl}A^b_m + \overset{2}{\Sigma} A^a_{ijk}A^b_{lm})\kappa^{ijklm}$$

$$+ (5\overset{2}{\Sigma} A^a_{ijklm}A^b_n + 10\overset{2}{\Sigma} A^a_{ijkmn}A^b_l + 8\overset{2}{\Sigma} A^a_{ijkm}A^b_{ln} + 6\overset{2}{\Sigma} A^a_{ijmn}A^b_{kl}$$

$$+ 3\overset{2}{\Sigma} A^a_{ijk}A^b_{lmn} + 9A^a_{ijm}A^b_{kln})\kappa^{ijkl}\kappa^{mn} + (10\overset{2}{\Sigma} A^a_{ijklm}A^b_n + 4\overset{2}{\Sigma} A^a_{ijkl}A^b_{mn}$$

$$+ 6\overset{2}{\Sigma} A^a_{ijlm}A^b_{kn} + 9A^a_{ijl}A^b_{kmn})\kappa^{ijk}\kappa^{lmn} + (60\overset{2}{\Sigma} A^a_{ijklmn}A^b_r$$

$$+ 45\overset{2}{\Sigma} A^a_{ijlmnr}A^b_k + 20\overset{2}{\Sigma} A^a_{ijkln}A^b_{mr} + 60\overset{2}{\Sigma} A^a_{ijlnn}A^b_{kr} + 15\overset{2}{\Sigma} A^a_{ilmnr}A^b_{jk}$$

$$+ 12\overset{2}{\Sigma} A^a_{ijkl}A^b_{mnr} + 18\overset{2}{\Sigma} A^a_{ijlm}A^b_{knr} + 36\overset{2}{\Sigma} A^a_{ijln}A^b_{kmr} + 36\overset{2}{\Sigma} A^a_{ilmn}A^b_{jkr})\kappa^{ijk}\kappa^{lm}\kappa^{nr}$$

$$+ (105\overset{2}{\Sigma} A^a_{ijklmnr}A^b_s + 90\overset{2}{\Sigma} A^a_{ijklmr}A^b_{ns} + 45\overset{2}{\Sigma} A^a_{ijklm}A^b_{nrs}$$

$$+ 60\overset{2}{\Sigma} A^a_{ijkmr}A^b_{lns} + 72A^a_{ijkm}A^b_{lnrs} + 24A^a_{ikmr}A^b_{jlns})\kappa^{ij}\kappa^{kl}\kappa^{mn}\kappa^{rs}\} + O(\nu^{-5}),$$

$$K^{abc} = {}_2\{A_i^a A_j^b A_k^c \kappa^{ijk} + 2 \overset{3}{\Sigma} A_{ik}^a A_j^b A_l^c \kappa^{ij}\kappa^{kl}\} + {}_3\{\overset{3}{\Sigma} A_{ij}^a A_k^b A_l^c \kappa^{ijkl}$$

$$+ (3 \overset{6}{\Sigma} A_{ijl}^a A_k^b A_m^c + 3 \overset{3}{\Sigma} A_{ilm}^a A_j^b A_k^c + 2 \overset{6}{\Sigma} A_{ij}^a A_{kl}^b A_m^c + 4 \overset{3}{\Sigma} A_{il}^a A_{jm}^b A_k^c)\kappa^{ijk}\kappa^{lm}$$

$$+ (12 \overset{3}{\Sigma} A_{ijkm}^a A_l^b A_n^c + 6 \overset{6}{\Sigma} A_{ijk}^a A_{lm}^b A_n^c + 6 \overset{6}{\Sigma} A_{ikm}^a A_{jl}^b A_n^c$$

$$+ 8 A_{ik}^a A_{jm}^b A_{ln}^c)\kappa^{ij}\kappa^{kl}\kappa^{mn}\} + {}_4\{(\overset{3}{\Sigma} A_{ijk}^a A_l^b A_m^c + \overset{3}{\Sigma} A_{ij}^a A_{kl}^b A_m^c)\kappa^{ijklm}$$

$$+ (4 \overset{6}{\Sigma} A_{ijkm}^a A_l^b A_n^c + 6 \overset{3}{\Sigma} A_{ijmn}^a A_k^b A_l^c + 2 \overset{6}{\Sigma} A_{ijk}^a A_{lm}^b A_n^c + 3 \overset{6}{\Sigma} A_{ijm}^a A_{kl}^b A_n^c$$

$$+ 6 \overset{6}{\Sigma} A_{ijm}^a A_{kn}^b A_l^c + 3 \overset{6}{\Sigma} A_{imn}^a A_{jk}^b A_l^c + 4 \overset{3}{\Sigma} A_{ij}^a A_{km}^b A_{ln}^c)\kappa^{ijkl}\kappa^{mn}$$

$$+ (4 \overset{3}{\Sigma} A_{ijkl}^a A_m^b A_n^c + 6 \overset{3}{\Sigma} A_{ijlm}^a A_k^b A_n^c + 6 \overset{6}{\Sigma} A_{ijl}^a A_{km}^b A_n^c + 3 \overset{6}{\Sigma} A_{ijl}^a A_{mn}^b A_k^c$$

$$+ 2 \overset{3}{\Sigma} A_{ij}^a A_{kl}^b A_{mn}^c + 4 A_{il}^a A_{jm}^b A_{kn}^c)\kappa^{ijk}\kappa^{lmn} + (20 \overset{3}{\Sigma} A_{ijkln}^a A_m^b A_r^c$$

$$+ 30 \overset{6}{\Sigma} A_{ijlmn}^a A_k^b A_r^c + 15 \overset{3}{\Sigma} A_{ilmnr}^a A_j^b A_k^c + 8 \overset{6}{\Sigma} A_{ijkl}^a A_{mn}^b A_r^c + 12 \overset{6}{\Sigma} A_{ijlm}^a A_{kn}^b A_r^c$$

$$+ 24 \overset{6}{\Sigma} A_{ijln}^a A_{km}^b A_r^c + 12 \overset{6}{\Sigma} A_{ijln}^a A_{mr}^b A_k^c + 12 \overset{3}{\Sigma} A_{ilmn}^a A_{jk}^b A_r^c + 24 \overset{6}{\Sigma} A_{ilmn}^a A_{jr}^b A_k^c$$

$$+ 18 \overset{6}{\Sigma} A_{ijl}^a A_{kmn}^b A_r^c + 9 \overset{6}{\Sigma} A_{ijl}^a A_{knr}^b A_m^c + 9 \overset{6}{\Sigma} A_{ijl}^a A_{mnr}^b A_k^c + 9 \overset{3}{\Sigma} A_{ilm}^a A_{jnr}^b A_k^c$$

$$+ 18 \overset{3}{\Sigma} A_{iln}^a A_{jmr}^b A_k^c + 12 \overset{6}{\Sigma} A_{ijl}^a A_{kn}^b A_{mr}^c + 12 \overset{3}{\Sigma} A_{ilm}^a A_{jn}^b A_{kr}^c + 6 \overset{6}{\Sigma} A_{iln}^a A_{jk}^b A_{mr}^c$$

$$+ 24 \overset{3}{\Sigma} A_{iln}^a A_{jm}^b A_{kr}^c + 6 \overset{6}{\Sigma} A_{imn}^a A_{ij}^b A_{kr}^c)\kappa^{ijk}\kappa^{lm}\kappa^{nr}$$

$$+ (90 \overset{3}{\Sigma} A_{ijklmr}^a A_n^b A_s^c + 30 \overset{6}{\Sigma} A_{ijklm}^a A_{nr}^b A_s^c + 60 \overset{6}{\Sigma} A_{ijkmr}^a A_{ln}^b A_s^c$$

$$+ 36 \overset{6}{\Sigma} A_{ijkm}^a A_{lnr}^b A_s^c + 36 \overset{6}{\Sigma} A_{ijkm}^a A_{lrs}^b A_n^c + 24 \overset{6}{\Sigma} A_{ikmr}^a A_{jln}^b A_s^c$$

$$+ 48 \overset{3}{\Sigma} A_{ijkm}^a A_{lr}^b A_{ns}^c + 24 \overset{3}{\Sigma} A_{ikmr}^a A_{jl}^b A_{ns}^c + 18 \overset{6}{\Sigma} A_{ijk}^a A_{lmr}^b A_{ns}^c$$

$$+ 18 \overset{3}{\Sigma} A_{ijk}^a A_{mnr}^b A_{ls}^c + 36 \overset{3}{\Sigma} A_{ikm}^a A_{jlr}^b A_{ns}^c)\kappa^{ij}\kappa^{kl}\kappa^{mn}\kappa^{rs}\} + O(\nu^{-5}),$$

$$K^{abcd} = {}_3\{A_i^a A_j^b A_k^c A_l^d \kappa^{ijkl} + 2 \overset{12}{\Sigma} A_{il}^a A_j^b A_k^c A_m^d \kappa^{ijk}\kappa^{lm} + (6 \overset{4}{\Sigma} A_{ikm}^a A_j^b A_l^c A_n^d$$

$$+ 4 \overset{12}{\Sigma} A_{ik}^a A_{jm}^b A_l^c A_n^d)\kappa^{ij}\kappa^{kl}\kappa^{mn}\} + {}_4\{\overset{4}{\Sigma} A_{ij}^a A_k^b A_l^c A_m^d \kappa^{ijklm} + (3 \overset{12}{\Sigma} A_{ijm}^a A_k^b A_l^c A_n^d$$

$$+ 3 \overset{4}{\Sigma} A_{imn}^a A_j^b A_k^c A_l^d + 2 \overset{24}{\Sigma} A_{ij}^a A_{km}^b A_n^c A_l^d + 4 \overset{6}{\Sigma} A_{im}^a A_{jn}^b A_k^c A_l^d)\kappa^{ijkl}\kappa^{mn}$$

$$+ (3 \overset{12}{\Sigma} A_{ijl}^a A_k^b A_m^c A_n^d + 2 \overset{12}{\Sigma} A_{ij}^a A_{kl}^b A_m^c A_n^d + 4 \overset{6}{\Sigma} A_{il}^a A_{jm}^b A_k^c A_n^d)\kappa^{ijk}\kappa^{lmn}$$

$$+ (12 \overset{12}{\Sigma} A_{ijln}^a A_k^b A_m^c A_r^d + 12 \overset{12}{\Sigma} A_{ilmn}^a A_j^b A_k^c A_r^d + 6 \overset{24}{\Sigma} A_{ijl}^a A_{kn}^b A_m^c A_r^d$$

$$+ 6 \overset{24}{\Sigma} A_{ijl}^a A_{mn}^b A_k^c A_r^d + 6 \overset{24}{\Sigma} A_{ilm}^a A_{jn}^b A_k^c A_r^d + 6 \overset{12}{\Sigma} A_{iln}^a A_{jk}^b A_m^c A_r^d$$

$$+ 12 \overset{24}{\Sigma} A_{iln}^a A_{jm}^b A_k^c A_r^d + 6 \overset{12}{\Sigma} A_{iln}^a A_{mr}^b A_j^c A_k^d + 6 \overset{12}{\Sigma} A_{lmn}^a A_{ir}^b A_j^c A_k^d$$

$$+4\overset{24}{\Sigma}A^a_{ij}A^b_{kl}A^c_{mn}A^d_r+8\overset{12}{\Sigma}A^a_{il}A^b_{jm}A^c_{kn}A^d_r+8\overset{12}{\Sigma}A^a_{il}A^b_{jn}A^c_{mr}A^d_k)\kappa^{ijk}\kappa^{lm}\kappa^{nr}$$

$$+(60\overset{4}{\Sigma}A^a_{ijkmr}A^b_jA^c_nA^d_s+24\overset{24}{\Sigma}A^a_{ijkm}A^b_{lr}A^c_nA^d_s+24\overset{12}{\Sigma}A^a_{ikmr}A^b_{jl}A^c_nA^d_s$$

$$+18\overset{12}{\Sigma}A^a_{ijk}A^b_{lmr}A^c_nA^d_s+18\overset{12}{\Sigma}A^a_{ikm}A^b_{jlr}A^c_nA^d_s+12\overset{24}{\Sigma}A^a_{ijk}A^b_{lm}A^c_{nr}A^d_s$$

$$+12\overset{24}{\Sigma}A^a_{ikm}A^b_{jl}A^c_{nr}A^d_s+24\overset{12}{\Sigma}A^a_{ikm}A^b_{jr}A^c_{ls}A^d_n$$

$$+16\overset{3}{\Sigma}A^a_{ik}A^b_{jm}A^c_{lr}A^d_{ns})\kappa^{ij}\kappa^{kl}\kappa^{mn}\kappa^{rs}\}+O(\nu^{-5}),$$

$$K^{abcde}={}_4\{A^a_iA^b_jA^c_kA^d_lA^e_m\kappa^{ijklm}+2\overset{20}{\Sigma}A^a_{im}A^b_jA^c_kA^d_lA^e_n\kappa^{ijkl}\kappa^{mn}$$

$$+2\overset{15}{\Sigma}A^a_{il}A^b_jA^c_kA^d_mA^e_n\kappa^{ijk}\kappa^{lmn}+(6\overset{30}{\Sigma}A^a_{iln}A^b_jA^c_kA^d_mA^e_r$$

$$+4\overset{60}{\Sigma}A^a_{il}A^b_{jn}A^c_kA^d_mA^e_r+4\overset{60}{\Sigma}A^a_{il}A^b_{mn}A^c_jA^d_kA^e_r)\kappa^{ijk}\kappa^{lm}\kappa^{nr}$$

$$+(24\overset{5}{\Sigma}A^a_{ikmr}A^b_jA^c_lA^d_nA^e_s+12\overset{60}{\Sigma}A^a_{ikm}A^b_{jr}A^c_lA^d_nA^e_s$$

$$8\overset{60}{\Sigma}A^a_{ik}A^b_{jm}A^c_{lr}A^d_nA^e_s)\kappa^{ij}\kappa^{kl}\kappa^{mn}\kappa^{rs}\}+O(\nu^{-5}),$$

$$K^{abcdef}={}_5\{A^a_iA^b_jA^c_kA^d_lA^e_mA^f_n\kappa^{ijklmn}+2\overset{30}{\Sigma}A^a_{in}A^b_jA^c_kA^d_lA^e_mA^f_r\kappa^{ijklm}\kappa^{nr}$$

$$+2\overset{60}{\Sigma}A^a_{im}A^b_jA^c_kA^d_lA^e_nA^f_r\kappa^{ijkl}\kappa^{mnr}+(6\overset{60}{\Sigma}A^a_{imr}A^b_jA^c_kA^d_lA^e_nA^f_s$$

$$+4\overset{180}{\Sigma}A^a_{im}A^b_{jr}A^c_kA^d_lA^e_nA^f_s+4\overset{120}{\Sigma}A^a_{im}A^b_{nr}A^c_jA^d_kA^e_lA^f_s)\kappa^{ijkl}\kappa^{mn}\kappa^{rs}$$

$$+(6\overset{90}{\Sigma}A^a_{ilr}A^b_jA^c_kA^d_mA^e_nA^f_s+4\overset{360}{\Sigma}A^a_{il}A^b_{jr}A^c_kA^d_mA^e_nA^f_s$$

$$+4\overset{90}{\Sigma}A^a_{ir}A^b_{ls}A^c_jA^d_kA^e_mA^f_n)\kappa^{ijk}\kappa^{lmn}\kappa^{rs}+(24\overset{60}{\Sigma}A^a_{ilns}A^b_jA^c_kA^d_mA^e_rA^f_t$$

$$+12\overset{360}{\Sigma}A^a_{iln}A^b_{js}A^c_kA^d_mA^e_rA^f_t+12\overset{360}{\Sigma}A^a_{iln}A^b_{ms}A^c_jA^d_kA^e_rA^f_t$$

$$+12\overset{180}{\Sigma}A^a_{lns}A^b_{im}A^c_jA^d_kA^e_rA^f_t+8\overset{120}{\Sigma}A^a_{il}A^b_{jn}A^c_{ks}A^d_mA^e_rA^f_t$$

$$+8\overset{720}{\Sigma}A^a_{il}A^b_{jn}A^c_{ms}A^d_kA^e_rA^f_t+8\overset{360}{\Sigma}A^a_{il}A^b_{mn}A^c_{rs}A^d_jA^e_kA^f_t)\kappa^{ijk}\kappa^{lm}\kappa^{nr}\kappa^{st}$$

$$+(120\overset{6}{\Sigma}A^a_{ikmrt}A^b_jA^c_lA^d_nA^e_sA^f_u+48\overset{120}{\Sigma}A^a_{ikmr}A^b_{jt}A^c_lA^d_nA^e_sA^f_u$$

$$+36\overset{90}{\Sigma}A^a_{ikm}A^b_{jrt}A^c_lA^d_nA^e_sA^f_u+24\overset{360}{\Sigma}A^a_{ikm}A^b_{jr}A^c_{lt}A^d_nA^e_sA^f_u$$

$$+24\overset{360}{\Sigma}A^a_{ikm}A^b_{jr}A^c_{st}A^d_lA^e_nA^f_u$$

$$+16\overset{360}{\Sigma}A^a_{ik}A^b_{jm}A^c_{lr}A^d_{nt}A^e_sA^f_u)\kappa^{ij}\kappa^{kl}\kappa^{mn}\kappa^{rs}\kappa^{tu}\}+O(\nu^{-6}).$$

In these expressions the summation signs refer to those permutations of a, b, ... which produce distinct terms (see above).

For the special case of a single function of a single variable, these formulae become

$$K_1 = {}_0\{A_0\} + {}_1\{A_2\kappa_2\} + {}_2\{A_3\kappa_3 + 3A_4\kappa_2^2\} + {}_3\{A_4\kappa_4 + 10A_5\kappa_3\kappa_2 + 15A_6\kappa_2^3\}$$
$$+ {}_4\{A_5\kappa_5 + 15A_6\kappa_4\kappa_2 + 10A_6\kappa_3^2 + 105A_7\kappa_3\kappa_2^2 + 105A_8\kappa_2^4\} + O(\nu^{-5}),$$

$$K_2 = {}_1\{A_1^2\kappa_2\} + {}_2\{2A_2A_1\kappa_3 + (6A_3A_1 + 2A_2^2)\kappa_2^2\} + {}_3\{(2A_3A_1 + A_2^2)\kappa_4$$
$$+ (20A_4A_1 + 18A_3A_2)\kappa_3\kappa_2 + (30A_5A_1 + 24A_4A_2 + 15A_3^2)\kappa_2^3\}$$
$$+ {}_4\{2A_4A_1 + 2A_3A_2)\kappa_5 + (30A_5A_1 + 28A_4A_2 + 15A_3^2)\kappa_4\kappa_2$$
$$+ (20A_5A_1 + 20A_4A_2 + 9A_3^2)\kappa_3^2 + (210A_6A_1 + 190A_5A_2 + 204A_4A_3)\kappa_3\kappa_2^2$$
$$+ (210A_7A_1 + 180A_6A_2 + 210A_5A_3 + 96A_4^2)\kappa_2^4\} + O(\nu^{-5}),$$

$$K_3 = {}_2\{A_1^3\kappa_3 + 6A_2A_1^2\kappa_2^2\} + {}_3\{3A_2A_1^2\kappa_4 + (27A_3A_1^2 + 24A_2^2A_1)\kappa_3\kappa_2$$
$$+ (36A_4A_1^2 + 72A_3A_2A_1 + 8A_2^3)\kappa_2^3\} + {}_4\{(3A_3A_1^2 + 3A_2^2A_1)\kappa_5$$
$$+ (42A_4A_1^2 + 84A_3A_2A_1 + 12A_2^3)\kappa_4\kappa_2 + (30A_4A_1^2 + 54A_3A_2A_1 + 10A_2^3)\kappa_3^2$$
$$+ (285A_5A_1^2 + 552A_4A_2A_1 + 297A_3^2A_1 + 252A_3A_2^2)\kappa_3\kappa_2^2 + (270A_6A_1^2 + 540A_5A_2A_1$$
$$+ 576A_4A_3A_1 + 216A_4A_2^2 + 270A_3^2A_2)\kappa_2^4\} + O(\nu^{-5}),$$

$$K_4 = {}_3\{A_1^4\kappa_4 + 24A_2A_1^3\kappa_3\kappa_2 + (24A_3A_1^3 + 48A_2^2A_1^2)\kappa_2^3\} + {}_4\{4A_2A_1^3\kappa_5$$
$$+ 48A_3A_1^3 + 72A_2^2A_1^2)\kappa_4\kappa_2 + (36A_3A_1^3 + 48A_2^2A_1^2)\kappa_3^2$$
$$+ (288A_4A_1^3 + 936A_3A_2A_1^2 + 288A_2^3A_1)\kappa_3\kappa_2^2 + (240A_5A_1^3 + 864A_4A_2A_1^2$$
$$+ 432A_3^2A_1^2 + 864A_3A_2^2A_1 + 48A_2^4)\kappa_2^4\} + O(\nu^{-5}),$$

$$K_5 = {}_4\{A_1^5\kappa_5 + 40A_2A_1^4\kappa_4\kappa_2 + 30A_2A_1^4\kappa_3^2 + (180A_3A_1^4 + 480A_2^2A_1^3)\kappa_3\kappa_2^2$$
$$+ (120A_4A_1^4 + 720A_3A_2A_1^3 + 480A_2^2A_1^2)\kappa_2^4\} + O(\nu^{-5}),$$

$$K_6 = {}_5\{A_1^6\kappa_6 + 60A_2A_1^5\kappa_5\kappa_2 + 120A_2A_1^5\kappa_4\kappa_3 + (360A_3A_1^5 + 1,200A_2^2A_1^4)\kappa_4\kappa_2^2$$
$$+ (540A_3A_1^5 + 1,800A_2^2A_1^4)\kappa_3^2\kappa_2 + (1,440A_4A_1^5 + 10,800A_3A_2A_1^4$$
$$+ 9,600A_2^3A_1^3)\kappa_3\kappa_2^3 + (720A_5A_1^5 + 5,760A_4A_2A_1^4 + 3,240A_3^2A_1^4$$
$$+ 17,280A_3A_2^2A_1^3 + 5,760A_2^4A_1^2)\kappa_2^5\} + O(\nu^{-6}).$$

REFERENCES

Fisher, R. A. (1930): Moments and product moments of sampling distributions. *Proc. London Math. Soc.*, Series 2, **30**, 199-238.

James, G. S. (1955): Cumulants of a transformed variate. *Biometrika*, **42**, 529-31.

———— (1958): On moments and cumulants of systems of statistics. *Sankhyā*, **20**, 1-30.

Kaplan, E. L. (1952): Tensor notation and the sampling cumulants of *k*-statistics. *Biometrika*, **39**, 319-23.

Paper received : September, 1960.

DISTRIBUTION OF THE DETERMINANT OF THE SUM OF PRODUCTS MATRIX IN THE NON-CENTRAL LINEAR CASE FOR SOME VALUES OF p

By O. P. BAGAI

Government College, Ludhiana, India

SUMMARY. The distribution of the determinant of a non-central Wishart matrix of order p with a non-centrality matrix of rank one is worked out for $p = 2, 3$ and 4.

1. GENERALIZED VARIANCE AND ITS MOMENTS

Wilks (1932) defines the generalized variance to be the determinant of the variance-covariance matrix and considers it to be a measure of the spread of the observations. Let S be the sample variance and covariance matrix with n degrees of freedom (d.f.) and $\Sigma(p \times p) = E(nS)$. The h-th moment of $|A| = |nS|$ in the central case is given by Wilks (1932). Let k_i^2 $(i = 1, 2, ..., p)$ be the positive roots of the determinantal equation

$$|T - k^2 \Sigma| = 0$$

where T is the non-centrality matrix of S. Assuming $k_i^2 = 0$ $(i = 2, 3, ..., p)$ and $k_1 \neq 0$, Anderson (1946) gives the h-th moment of $|A|$ in the non-central case as follows :

$$E|A|^h = 2^{ph} \exp\left(-\tfrac{1}{2}k_1^2\right) \prod_{i=1}^{p-1} \frac{\Gamma\left(\frac{n-i}{2}+h\right)}{\Gamma\left(\frac{n-i}{2}\right)} \cdot \sum_{j=0}^{\infty} \frac{k_1^{2j}}{2^j j!} \cdot \frac{\Gamma\left(\frac{n}{2}+j+h\right)}{\Gamma\left(\frac{n}{2}+j\right)} \quad \dots \quad (1.1)$$

Making use of this moment, we find below the distribution of $|A|$ in the non-central linear case for p equal to 2, 3, and 4.

2. SOME PRELIMINARIES AND INTEGRALS

We list below Legendre's duplication formula and the values of some definite integrals obtained from standard books of tables on integrals. Also we list two other definite integrals which we have ourselves evaluated and published elsewhere.

(i) Legendre's duplication formula for the gamma function :

$$\Gamma\left(n+\tfrac{1}{2}\right) \Gamma(n+1) = \frac{\pi^{\frac{1}{2}}\Gamma(2n+1)}{2^{2n}} \quad \dots \quad (2.1)$$

(ii) For $a \geqslant 0$, Larsen's Table (1948) gives

$$\int_0^\infty \exp\left[-(x^2+a^2x^{-2})\right] dx = \tfrac{1}{2}\sqrt{\pi} \exp(-2a) \quad \dots \quad (2.2)$$

(iii) In Table 98 (pp. 143-144) Bierens de Hann gives the following two integrals:*

$$\int_0^\infty x^{a-\frac{1}{2}} \exp\left[-(px+qx^{-1})\right] dx$$

$$= (q/p)^{\frac{1}{2}a} \exp\left(-2\sqrt{pq}\right) \frac{\sqrt{\pi}}{p} \sum_{n=0}^\infty \frac{(a+1-n)^{2n/1}}{2^{\frac{1}{2}n} (2\sqrt{pq})^n} \qquad \cdots \quad (2.3)$$

and

$$\int_0^\infty x^{-a-\frac{1}{2}} \exp\left[-(px+qx^{-1})\right] dx$$

$$= (p/q)^{\frac{1}{2}a} \exp\left(-2\sqrt{pq}\right) \frac{\sqrt{\pi}}{p} \sum_{n=0}^\infty \frac{(a-n)^{2n/1}}{2^{\frac{1}{2}n} (2\sqrt{pq})^n} \qquad \cdots \quad (2.4)$$

(iv) We now give two other integrals which we have evaluated of our own (published elsewhere). Their evaluation is outlined as follows :

(a) Consider $I = \int_0^\infty x \exp\left[-2(x+ax^{-1})\right] dx$.

It can be easily evaluated by setting $a = \frac{3}{2}, p = 2, q = 2a$ in (2.3); but we have evaluated it by a different method as follows :

Set $x = \frac{1}{2} u^{-1}$ and $b = 4a$. We obtain $K(b) = \frac{1}{4} \int_0^\infty u^{-3} \exp\left(-u^{-1}-bu\right) dy$ which satisfies the following differential equation :

$$b \frac{d^2K}{db^2} - \frac{dK}{db} - K = 0.$$

Solving this differential equation, we set $b = 4a$ and obtain :

$$\int_0^\infty x \exp\left[-2(x+ax^{-1})\right] dx = \left[\frac{-(1+2\gamma)-\log 4a}{4} \right.$$

$$\sum_{r=0}^\infty \frac{(4a)^{r+2}}{r!\,(r+2)!} + \frac{1}{4} \left\{ 1-(4a)+\frac{(4a)^2}{2^2}+\frac{11}{2^2}\frac{(4a)^3}{3^2}+\cdots \right\} \qquad \cdots \quad (2.5)$$

where γ is Euler's constant.

(b) Consider the integral

$$L_r(a) = 2 \int_0^\infty x^{2r+1} \exp\left(-x^2-ax^{-1}\right) dx \qquad \cdots \quad (2.6)$$

where a is real and positive.

It satisfies the differential equation

$$a \frac{d^3L_r}{da^3} - 2r \frac{d^2L_r}{da^2} + 2L_r = 0. \qquad \cdots \quad (2.7)$$

---

* In both of these, Kramp's notation is used, namely, $x^{n/h} = x(x+h)(x+2h)\ldots(x+n-1h)$.

Solving this equation, we have

$$L_0(a) = \left(\gamma - \log a \right) \left[a^2 - 2 \sum_{r=1}^{\infty} (-1)^{r+1} \frac{a^{2r+2}}{r!(2r+2)!} \right]$$

$$+ \left[1 + \frac{2.3}{2^2.1^2} a^2 - \frac{4(124)}{4^2.3^2.2^2.1^2} a^4 + \dots \right]$$

$$- \Gamma(\tfrac{1}{2}) \left[a - \sum_{r=1}^{\infty} (-1)^{r+1} \frac{(r-1)! \, 2^{2r-1}}{(2r-1)! \, (2r+1)!} a^{2r+1} \right] \qquad \dots \quad (2.8)$$

$$L_1(a) = 2[(\gamma + \tfrac{1}{2}) - \log a] \left[\frac{a^4}{4!} - \sum_{r=1}^{\infty} (-1)^{r+1} \frac{a^{2r+4}}{r!(2r+4)!} \right]$$

$$+ \left[1 + \tfrac{1}{2}a^2 + \frac{19}{144} a^4 + \dots \right]$$

$$- \Gamma(\tfrac{3}{2}) \left(a + \sum (-1)^{r+1} \frac{2^r a^{2r+1}}{(2r+1)!} \dots \right) \qquad \dots \quad (2.9)$$

$$L_2(a) = [(\gamma + \tfrac{5}{2}) - \Gamma(3) \log a] \left[\frac{a^6}{6!} - \sum_{r=1}^{\infty} (-1)^{r+1} \frac{a^{2r+6}}{r!(2r+6)!} \right]$$

$$+ \Gamma(3) \left(1 + \frac{1}{4} a^2 + \frac{1}{48} a^4 + \frac{17}{10} \frac{a^6}{61} + \dots \right)$$

$$- \Gamma(\tfrac{5}{2}) \left(a + \frac{2}{1.3} \frac{a^3}{3!} + \frac{2^2}{1.3} \frac{a^5}{5!} + \frac{2^3}{1.3} \frac{a^7}{7!} \right.$$

$$\left. + \frac{2^4}{1.3} \frac{1}{1.3} \frac{a^9}{9!} + \frac{2^5}{1.3} \frac{1}{1.3.5} \frac{a^{11}}{11!} + \dots \right). \qquad \dots \quad (2.10)$$

$L_r(a)$ for $r = 3, 4, \dots$ can be evaluated the same way as above by obtaining the solutions of their respective differential equations.

3. DERIVATION OF THE SUITABLE FORM FOR FINDING THE DISTRIBUTION OF $|A|$

From (1.1) it is clear that

$$E|A|^h = E(u_0 u_1 \dots u_{r-1})^h,$$

if the joint density of the u_i 's is

$$\prod_{i=1}^{p-1} \left\{ \frac{1}{\Gamma\left[\frac{n-i}{2}\right]} \left[\frac{u_i}{2} \right]^{\frac{n-i}{2}-1} \exp \left(-\tfrac{1}{2} u_i \right) d \left[\frac{u_i}{2} \right] \right\}$$

$$\frac{\exp \left(-\tfrac{1}{2} k_i^2 \right)}{\Gamma\left[\frac{n}{2}\right]} (\tfrac{1}{2} u_0)^{\frac{1}{2}n-1} e^{-\frac{1}{2} u_0} \sum_{r=0}^{\infty} \frac{\Gamma(\tfrac{1}{2} n)}{\Gamma(\tfrac{1}{2} n + r)} \cdot \frac{(k_1^2 u_0 / 4)^r}{r!} - d(\tfrac{1}{2} u_0)$$

$$(0 \leqslant u_i < \infty), \qquad (i = 0, 1, 2, \dots, \overline{p-1}).$$

57

8

From this we calculate that the distribution of $|A|$ is the same as that of $u_0, u_1, \ldots u_{p-1}$. After a little manipulation and setting $n = 2m+p+1$ $(p \leqslant n)$, the joint distribution of u_i $(i = 0, 1, 2, \ldots, p-1)$ becomes

$$\frac{1}{2^{p(m+\frac{1}{2}p+\frac{1}{2})}} \prod_{r=0}^{p-1} \frac{u_{p-r-1}^{m+\frac{1}{2}r}}{\Gamma(m+\frac{1}{2}r+1)} \exp\left[-\frac{1}{2}\sum_{i=1}^{p-1} u_i\right] \exp\left(-\frac{1}{2}k_1^2\right)$$

$$\left[1+\frac{u_0}{1!}\frac{(k_1^2/2)}{2m+p+1} + \frac{u_0^2}{2!}\frac{(k_1^2/2)^2}{(2m+p+1)(2m+p+3)}+\ldots\right] \prod_{i=0}^{p-1} du_i \quad \ldots \ (3.1)$$

$$(0 \leqslant u_1 < \infty), \quad (i = 1, \ldots, p-1).$$

4. DISTRIBUTION OF $|A|$ UP TO THE ORDER 4 IN THE NON-CENTRAL LINEAR CASE

Case 1 : For $p = 2$, the joint distribution of u_0 and u_1 from (3, 4) is

$$\frac{2^{-2(m+\frac{1}{2})} u_1^m u_0^{m+\frac{1}{2}}}{\Gamma(m+1)\,\Gamma(m+\frac{1}{2})} \exp\{-\frac{1}{2}(u_0+u_1+k_1^2)\}$$

$$\left[1+\frac{u_0}{1!}\frac{(k_1^2/2)}{2m+3} + \frac{u_0^2}{2!}\frac{(k_1^2/2)^2}{(2m+3)(2m+5)} + \ldots\right] du_0\,du_1, \quad \ldots \ (4.1)$$

$$(0 \leqslant u_0, u_1 < \infty).$$

Making use of (2.1) and setting $u_0 u_1 = V_1^2$, $u_0 = 2V_2^2$, we get from (4.1) the simultaneous distribution of V_1 and V_2 as follows :

$$\frac{2}{\sqrt{\pi}}\frac{V_1^{2m+1}}{\Gamma(2m+2)} \exp\left(-\frac{1}{2}k_1^2\right)\exp\left(-\frac{V_1^2}{4V_2^2}-V_2^2\right)$$

$$\left[1+\frac{V_2^2}{1!}\frac{k_1^2}{2m+3} + \frac{V_2^4}{2!}\frac{k_1^4}{(2m+2)(2m+5)}+\ldots\right]dV_1 dV_2$$

$$(0 < V_1, V_2 < \infty).$$

The distribution of $V_1(=\sqrt{u_0 u_1})$ is then

$$\frac{2}{\sqrt{\pi}}\frac{\exp\left(-\frac{1}{2}k_1^2\right)V_1^{2m+1}dV_1}{\Gamma(2m+2)} \int_0^\infty \exp\left(-\frac{V_1^2}{4V_2^2}-V_2^2\right)$$

$$\left(1+\frac{V_2^2}{1!}\frac{k_1^2}{2m+3} + \frac{V_2^4}{2!}\frac{k_1^4}{(2m+3)(2m+5)}+\ldots\right)dV_2 \quad \ldots \ (4.2)$$

$$(0 \leqslant V_1 < \infty).$$

Now consider, $$I_r = \int_0^\infty V_2^{2r} \exp\left\{-\frac{V_1^2}{4V_2^2}-V_2^2\right\}dV_2.$$

For $r = 0$, we use (4.2) and obtain,

$$\int_0^\infty \exp\left\{-\frac{V_1^2}{4V_2^2}-V_2^2\right\}dV_2 = \frac{\sqrt{\pi}}{2}\exp(-V_1) \qquad \ldots \quad (4.3)$$

and for $r \neq 0$ we set $V_2^2 = t$ and the integral I_r reduces to

$$I_r = \tfrac{1}{2}\int_0^\infty t^{r-1}\exp\left(-t-\frac{V_1^2}{4t}\right)dt.$$

We now use (2.3) and obtain for $(r \neq 0)$

$$I_r = \tfrac{1}{2}\left(\frac{V_1}{2}\right)^r \exp(-V_1)\sqrt{\pi}\sum_{n=0}^\infty \left[\frac{(r+1-n)^{2n/1}}{2^{in}\,V_1^n}\right]$$

or $\qquad\qquad I_r = \dfrac{\sqrt{\pi}}{2}\exp(-V_1)T_r \qquad\qquad \ldots \quad (4.4)$

where $\qquad\qquad T_r = \left[\dfrac{V_1}{2}\right]^r \sum_{n=0}^\infty \left[\dfrac{(r+1-n)^{2n/1}}{2^{in}V_1^n}\right]. \qquad \ldots \quad (4.5)$

Thus, with the help of (4.3) and (4.4), we obtain the distribution of $V_1 (= \sqrt{u_0\,u_1})$ from (4.2) as follows :-

$$\frac{V_1^{2m+1}\exp(-V_1-\tfrac{1}{2}k_1^2)}{\Gamma(2m+2)}\left[1+\frac{T_1}{1!}\frac{k_1^2}{2m+3}\right.$$

$$\left.+\frac{T_2}{2!}\frac{k_1^4}{(2m+3)(2m+5)}+\ldots\right]dV_1 \qquad \ldots \quad (4.6)$$

where $\quad 0 \leqslant V_1 < \infty, \quad m = \dfrac{n-3}{2}$ and $T_r\,(r \neq 0)$ is defined by (4.5).

It follows from (4.6) that in the central case, i.e., when $k_1^2 = 0$ and $m = \tfrac{1}{2}(n-3)$, the distribution of $V_1 = \sqrt{u_0\,u_1}$ is

$$\frac{1}{\Gamma(n-1)}\;V_1^{n-2}\exp(-V_1)dV_1 \quad \text{for} \quad 0 \leqslant V_1 < \infty \qquad \ldots \quad (4.7)$$

which is that of gamma variate with parameter $(n-1)$.

Case 2 : For $p = 3$, the joint distribution of u_0, u_1 and u_2 from (3.1) can be written as follows :

$$\frac{2^{-3(m+\frac{1}{2})}\,u_2^m\,u_1^{m+\frac{1}{2}}\,u_0^{m+1}}{\Gamma(m+1)\,\Gamma(m+\frac{3}{2})\,\Gamma(m+2)}\exp\left[-\tfrac{1}{2}\sum_{i=0}^\infty u_i-\tfrac{1}{2}k_1^2\right]$$

$$\times\left[1+\frac{u_0}{1!}\frac{k_1^2/2}{2m+4}+\frac{u_0^2}{2!}\frac{(k_1^2/2)^2}{(2m+4)(2m+6)}+\ldots\right]du_0du_1du_2 \qquad \ldots \quad (4.8)$$

where $\qquad 0 \leqslant u_0, u_1, u_2 < \infty.$

Using (2.1) and setting $u_0 u_1 u_2 = V_1$, $u_1 u_0^2 = V_2^2$, $u_0 = 2V_3^2$ in (4.8) we obtain the distribution of V_1 after a little manipulation as follows :

$$\frac{1}{\sqrt{\pi} 2^{m+4}} \frac{V_1^m \exp\left(-\frac{1}{2}k_1^2\right)}{\Gamma(m+1)\Gamma(2m+3)} \int_{V_3=0}^{\infty} \int_{V_2=0}^{\infty} V_3^{-1} \exp\left(-\frac{V_1 V_3^2}{V_2^2} - \frac{V_2^2}{8V_3^4} - V_3^2\right)$$

$$\left[1 + \frac{V_3^2}{1!} \frac{k_1^2}{2m+4} + \frac{V_3^4}{2!} \frac{k_1^4}{(2m+4)(2m+6)} + \dots\right] dV_3 \, dV_2 \, dV_1 \qquad \dots \text{(4.9)}$$

where $0 \leqslant V_1 < \infty$.

Using (4.2), $\quad \int_0^\infty \exp\left[-\frac{V_1 V_3^2}{V_2^2} - \frac{V_2^2}{8V_3^4}\right] dV_2 = \sqrt{2\pi}\, V_3^2 \exp\left[-\frac{1}{V_3}\sqrt{\frac{V_1}{2}}\right].$

Then (4.9) reduces to

$$\frac{V_1^m \exp\left(-\frac{k_1^2}{2}\right)}{2^m \Gamma(m+1)\,\Gamma(2m+3)} \int_{V_3=0}^{\infty} \exp\left(-\frac{1}{V_3}\sqrt{\frac{V_1}{2}} - V_3^2\right)$$

$$\left[V_3 + \frac{V_3^3}{1!} \frac{k_1^2}{2m+4} + \frac{V_3^5}{2!} \frac{k_1^4}{(2m+4)(2m+6)} + \dots\right] dV_1 \, dV_3 \qquad \dots \text{(4.10)}$$

where $0 \leqslant V_1 < \infty$.

Now making use of the integral (2.6) for $r = 0, 1, 2 \dots$ given respectively in (2.8), (2.9) and (2.10) etc, and remembering that a in (2.6) is equal to $(\frac{1}{2}V_1)^{\frac{1}{2}}$, the distribution of $V_1 (= u_0 u_1 u_2)$ is

$$\frac{V_1^m \exp\left(-\frac{1}{2}k_1^2\right)}{2^{m+1}\Gamma(m+1)\Gamma(2m+3)} \left[L_0\left(\sqrt{\frac{V_1}{2}}\right) + \frac{k_1^2}{1!} \frac{L_1\left(\sqrt{\frac{V_1}{2}}\right)}{2m+4}\right.$$

$$\left. + \frac{k_1^4}{2!} \frac{L_2\left(\sqrt{\frac{V_1}{2}}\right)}{(2m+4)(2m+6)} + \dots\right] dV \qquad \dots \text{(4.11)}$$

where $0 \leqslant V_1 < \infty$ and $m = (n-4)/2$.

For the central case, i.e., when $k_1^2 = 0$ and $m = (n-4)/2$, (4.11) becomes

$$\frac{V_1^{(n-4)/2}}{2^{(n-2)/2}\Gamma\left(\frac{n}{2} - 1\right)\Gamma(n-1)} L_0\left(\sqrt{\frac{1}{2}V_1}\right) \qquad (0 \leqslant V_1 < \infty) \qquad \dots \text{(4.12)}$$

where $L_0\left(\sqrt{\frac{1}{2}V_1}\right)$ is defined by (2.8).

Case 3: For $p = 4$, the joint distribution of u_0, u_1, u_2 and u_3 from (3.4) can be written as follows :-

$$\frac{2^{-4(m+\frac{7}{4})}\, u_3^m\, u_2^{m+\frac{1}{2}}\, u_1^{m+1}\, u_0^{m+\frac{3}{2}}}{\Gamma(m+1)\,\Gamma(m+\frac{3}{2})\,\Gamma(m+2)\,\Gamma(m+\frac{5}{2})} \;\exp\left(-\tfrac{1}{2}\sum_{i=0}^{3} u_i - \tfrac{1}{2}k_1^2\right)$$

$$\left[1+\frac{u_0}{1!}\frac{(k_1^2/2)}{2m+5}+\frac{u_0^2}{2!}\frac{(k_1^2/2)^2}{(2m+5)(2m+7)}+\ldots\right] du_0\, du_1\, du_2\, du_3 \qquad \ldots \quad (4.13)$$

$$(0 \leqslant u_0, u_1, u_2, u_3 < \infty).$$

Using (2.1) and setting $u_3 u_2 u_1 u_0 = V_1$, $u_2 u_1 u_0 = 2V_2^2$, $u_1 u_0 = V_3^2$, $u_0 = 2V_4^2$ in (4.13), we obtain the distribution of $V_1(=u_0 u_1 u_2 u_3)$ after a little manipulation as follows :

$$\frac{2V_1^m \exp(-\tfrac{1}{2}k_1^2)}{\sqrt{\pi}\,\Gamma(2m+2)\Gamma(2m+4)} \int_{V_4=0}^{\infty}\int_{V_3=0}^{\infty}\int_{V_2=0}^{\infty} \exp\left(-\frac{V_1}{4V_2^2}-\frac{V_2^2}{V_3^2}-\frac{V_3^2}{4V_4^2}-V_4^2\right)$$

$$\left[1+\frac{V_4^2}{1!}\frac{k_1^2}{2m+5}+\frac{V_4^4}{2!}\frac{k_1^4}{(2m+5)(2m+7)}+\ldots\right] dV_1\, dV_2\, dV_3\, dV_4 \qquad \ldots \quad (4.14)$$

where $0 \leqslant V_1 < \infty$.

Making use of (2.2), we integrate (4.14) with respect to V_2 and obtain,

$$\frac{V_1^m \exp(-\tfrac{1}{2}k_1^2)}{\sqrt{\pi}\,\Gamma(2m+2)\,\Gamma(2m+4)} \int_{V_3=0}^{\infty} V_3 \exp\left(-\frac{\sqrt{V_1}}{V_3}\right) \int_{V_4=0}^{\infty}\exp\left(-\frac{V_3^2}{4V_4^2}-V_4^2\right)$$

$$\left[1+\frac{V_4^2}{1!}\frac{k_1^2}{2m+5}+\frac{V_4^4}{2!}\frac{k_1^4}{(2m+5)(2m+7)}+\ldots\right] dV_1\, dV_3\, dV_4 \quad \ldots \quad (4.15)$$

where $0 \leqslant V_1 < \infty$.

To integrate with respect to V_4, we evaluate again the first integral as before by using (2.2), while in the others we set $V_4^2 = t$ and then, using (2.3), we obtain in place of (4.15) the following :

$$\frac{V_1^m \exp(-\tfrac{1}{2}k_1^2)}{2\Gamma(2m+2)\,\Gamma(2m+4)} \int_{V_3=0}^{\infty} V_3 \exp\left(-\frac{\sqrt{V_1}}{V_3}-V_3\right)$$

$$\left[1+\frac{I_1'}{1!}\frac{k_1^2}{2m+5}+\frac{I_2'}{2!}\frac{k_1^4}{(2m+5)(2m+7)}+\ldots\right] dV_3 \qquad \ldots \quad (4.16)$$

where $0 \leqslant V_1 < \infty$

and $\qquad\qquad I_r' = \tfrac{1}{2}\left(\frac{V_3}{2}\right)^r \sqrt{\pi}\exp(-V_3)\sum_{n=0}^{\infty}\frac{(r+1-n)^{2n/1}}{2^{n/2}V_3^n}.$ $\qquad \ldots \quad (4.17)$

Further to evaluate (4.16), we have to use either (2.3) or (2.4) for $p = 1$, $q = \sqrt{V_1}$ and various suitable values of a. This determines the distribution of $V_1 (= u_0 u_1 u_2 u_3)$ where it should be remembered that $m = \frac{1}{2}(n-5)$.

For the central case, we set $k_1^2 = 0$ in (4.16), and we obtain the distribution of V_1 as follows :

$$\frac{V_1^m \exp\left(-\frac{1}{2}k_1^2\right)}{2\Gamma(2m+2)\,\Gamma(2m+4)} \int_{V_3=0}^{\infty} V_3 \exp\left[-\frac{\sqrt{V_1}}{V_3} - V_3\right] dV_3\, dV_1 \;(0 \leqslant V_1 < \infty) \quad \dots \quad (4.18)$$

To evaluate (4.18), we can, of course, make use of (2.3) for $a = \frac{3}{2}$, $p = 1$, $q = \sqrt{V_1}$, but we prefer to use (2.5) and then write the distribution of $V_1 = u_0 u_1 u_2 u_3$ in the central case as follows :

$$\frac{V_1^{\frac{n-5}{2}}\, dV_1}{\Gamma(n-3)\,\Gamma(n-1)} \left[\frac{(1+2\gamma)-\log a}{2}\left(\frac{a^2}{2!0!} + \frac{a^3}{3!1!} + \frac{a^4}{4!2!} + \dots\right) \right.$$

$$\left. + \frac{1}{2}\left(1 - a + \frac{1}{2!}a^2 + \frac{11}{3^2\,2^2}a^3 + \dots\right)\right] \quad \dots \quad (4.19)$$

where $0 \leqslant V_1 < \infty$ and $a = \sqrt{V_1}$.

We could also use, by setting $r = 2$, $p = 1$, $\alpha = 4\sqrt{v_1}$, in the following integral :

$$\int_0^{\infty} t^{r-1}\, e^{-pt-\alpha/4t}\, dt = 2\left(\frac{\alpha}{4t}\right)^{\frac{1}{2}r} K_r(\sqrt{\alpha\, p}) \;\text{(Bateman, p. 146 (29))}$$

where $K_r(z)$ denotes the Modified Bessel Function.

REFERENCES

Anderson, T. W. (1946): The noncentral Wishart distribution and certain problems of multivariate statistics. *Ann. Math. Stat.*, **17**, 409-431.

Bierens de Haan, D. (1939): *Nonvelles Tables d' Integrates Definies*, Ed. of 1867 corrected. G. E. Stechert and Co., New York.

Larsen, H. D. (1948): *Rinehart Mathematical Tables, Formulas and curves*, Rinehart and Co. Inc., New York.

Wilks, S. S. (1932): Certain generalizations in the analysis of variance. *Biometrika*, **26**, 471-494.

Paper received: November, 1961.

POLYSOMIC INHERITANCE AND THE THEORY OF SHUFFLING

By P. A. P. MORAN

Australian National University, Canberra

SUMMARY. . A theory of shuffling is applied to elucidate difference between chromosomal and chromatid segregation in polysomic inheritance. Attention is drawn to unsolved problems in shuffling theory.

In diploid individuals the chromosomes (other than the sex chromosomes) occur in similar pairs and the process of segregation, considered for factors at a single locus, is quite simple in its end result. Thus, if there are two possible factors, A and B, at a single locus the gametes produced by an individual of constitution AB will be A or B with probabilities $\frac{1}{2}$ wherever the locus is situated on the chromosomes. In actual fact the cytological process which brings this about is considerably more complicated than might appear from the above statement.

The situation in autopolyploids is quite different and much more complicated. The number of chromosomes in each individual zygote may be even or odd. We confine the discussion in this paper to the case where the number is even and the more detailed algebra to the case where there are four chromosomes (tetraploids).

Segregation can then occur in more than one way. To illustrate this suppose that there are only two possible alleles, A and B, at a single locus so that a gamete or zygote may be written $A^x B^y$ where $x + y = s$. For tetraploids we will have $s = 2$ for a gamete and $s = 4$ for a zygote.

If we have a zygote of form $A^r B^{2m-r}$ its offspring gametes will be of the form $A^s B^{m-s}$ but there is more than one way of determining the probabilities of production of the various types of gamete from a specified zygote. Two cases most frequently considered are known as chromosomal and chromatid segregation and it appears that most observed cases are either one of these or a mixture of the two.

In chromosomal segregation the gamete is formed by a process which is equivalent to choosing m chromosomes at random, and without replacement, from the set represented by $A^r B^{2m-r}$. Thus the probability of obtaining a gamete of the form $A^s B^{m-s}$ is

$$\binom{2m}{m}^{-1} \binom{r}{s} \binom{2m-r}{m-s}$$

In chromatid segregation, on the other hand, the gamete, containing m chromosomes, is formed by a process which is equivalent to assuming that each of the $2m$ chromosomes in the parent zygote divide into two thus forming a set which we may

denote as $A^{2r}B^{4m-2r}$, and that from these a selection of m chromosomes is chosen at random without replacement. Thus the probability of obtaining a gamete of form A^sB^{m-s} is

$$\binom{4m}{m}^{-1}\binom{2r}{s}\binom{4m-2r}{m-s}$$

This results in quite a different set of probabilities and in particular makes possible the formation of gametes of types which could not be formed by chromosomal segregation. For example a zygote of form AB^3 will give gametes of form AB and B^2 with probabilities $\frac{1}{2}$ under chromosomal segregation and gametes of form AA, AB and BB with probabilities $\frac{1}{28}, \frac{3}{7}, \frac{15}{28}$ with chromatid segregation.

Fisher and Mather (1943) have introduced parameters to describe types intermediate between chromosomal and chromatid segregation. Consider a tetraploid zygote. Its diploid gametic offspring might be formed of two chromosomes identical with two different chromosomes chosen at random from the four chromosomes of its parent. This is what always happens in chromosomal segregation but only happens with probability $\frac{6}{7}$ in chromatid segregation. Alternatively, the two chromosomes in the gamete might be identical with a single chromosome of the parent, chosen at random. This never happens in chromosomal segregation but occurs with probability $\frac{1}{7}$ in chromatid segregation. Cases intermediate between the chromosomal and chromatid segregation can be described by writing α for the probability of the second mode of formation. Then a zygote of form AB^3 will produce gametes of forms AA, AB and BB with probabilities $\frac{1}{2}\alpha$, $\frac{1}{2}(1-\alpha)$ and $\frac{1}{2}+\frac{1}{2}\alpha$. The parameter α thus completely specifies the mode of segregation. With hexaploids, the triploid gamete may be derived from three different chromosomes with probability, $1-\beta$ say, or from two only (one supplying two chromatids) with probability β. Again, a single parameter, β, completely specifies the mode of segregation but with octaploids two parameters are necessary.

In general, chromosomal segregation occurs for loci which are near the centromere and chromatid segregation for loci which are far from it. The problem we have to consider is to explain how this comes about in terms of the actual process of segregation.

The gamete contains only one half of the number of chromosomes in the zygote but it is not produced by a simple process of reduction or halving but by a process which involves first a duplication and then two reductions.

Suppose that there are $2s$ chromosomes. Meiosis begins with a pairing of these chromosomes along their length thus forming s pairs but this pairing does not remain the same along the whole length. Thus if there are four chromosomes denoted by 1, 2, 3 and 4, we might have a pairing such as (12) (34) near the centromere (the part of the chromosome which controls the ultimate splitting) and at some distance along each arm this may change to (13)(24) or (14)(23) only to change again further

along. In a tetraploid the resulting figure is very like that formed by the four chromatids in a diploid just after crossing over with the difference that the four centromeres are here distinct.

The chromosomes then each split into two chromatids so that there are now $4s$ chromatids in all. At any place where two chromosomes were paired there are now four chromatids and between these crossing-over occurs. The essential feature of the situation is that owing to the interchanges in partners in the pairing of the chromosomes crossing-over may not only occur between, say, chromosomes 1 and 2, but also between 1 and 3, 1 and 4 and so on, at points where these chromosomes are paired. The end result after a sufficient amount of re-pairing and crossing-over has taken is that the $4s$ elements of the chromatids at a locus far from the centromere have been shuffled about so thoroughly that they are attached to the $4s$ portions of chromatid near the centromere in a manner which is effectively random, i.e. they have been "shuffled" into a random permutation. The aim of the mathematical theory is to give a description of this process of shuffling.

The cell then undergoes divisions twice giving four cells each containing s chromatids which become the chromosomes of the gamete. In the first division the $2s$ pairs of chromatids are separated into two sets of s pairs in a completely random manner ("random disjunction"). At the second division the two chromatids in each pair come apart, beginning at the centromeres, and each enters a different cell. Thus we see that near the centromere where there is no re-pairing and crossing-over between the locus and the centromere, segregation will be chromosomal whilst a long way away from the centromere random shuffling will result in chromatid segregation.

We now confine ourselves to tetraploids. If there are only two alleles, A and B say, the forms of zygote we have to consider are A^3B and A^2B^2 (AB^3 having the same theory as A^3B). The theory is then particularly simple. However, we consider here the general case of four alleles so that the zygote is represented by the symbol $ABCD$. We now introduce a symbol to denote the manner in which the eight elements $AABBCCDD$ representing the alleles at the locus considered on the eight chromatids are joined to the four centromeres which are associated in pairs. Thus if no re-pairing or crossing-over has taken place the initial state can be described by the symbol

$$\{[(AA)(BB)][(CC)(DD)]\}$$

or by one of the similar symbols obtained by permuting the letters A, B, C and D. Thus the above symbol indicates that the two chromatids carrying the allele A are joined to the same centromere which is paired with a centromere joined to the two chromatids carrying B, and similarly for C and D. This symbol can also represent a state in which re-pairing and crossing-over has occurred in such a way that as far as the locus under consideration is concerned the joining with the centromeres is as described. We denote by S_1 this state and all states obtained from it by permutation of the letters A, B, C and D. We shall also use S_1 to denote the sum of the probabilities of this state and its permutations, and since the initial pairing at the centromeres is random, all such permutations are equally probable.

9

We now have to enumerate all possible configurations and in order to do so it is a help to classify them. This can be done in two ways. Let n be the total number of doublets, i.e. symbols of the form (AA), (BB), etc. Then n can take the values 0, 1, 2 and 4. Similarly let m be the total number of different alleles which are joined to one centromere pair. This is equal to the total number joined to the other centromere pair. m can take the values 2, 3 and 4, but not all pairs of values of m and n are possible. Thus if $m = 4$, n must equal 0. The number of possible pairs (m, n) is seven but to one of these corresponds two configurations. Table 1 shows the possible configurations with typical symbols, all symbols obtained by permutation of A, B, C and D being equiprobable.

TABLE 1.   THE EIGHT POSSIBLE CONFIGURATIONS FOR
A TETRAPLOID

	configuration	n	m
S_1	$[(AA)(BB)][(CC)(DD)]$	4	2
S_2	$[(AB)(AB)][(CC)(DD)]$	2	2
S_3	$[(AB)(CC)][(AB)(DD)]$	2	3
S_4	$[(AA)(BC)][(DB)(DC)]$	1	3
S_5	$[(AB)(AB)][(CD)(CD)]$	0	2
S_6	$[(AB)(AC)][(BD)(CD)]$	0	3
S_7	$[(AB)(CD)][(AB)(CD)]$	0	4
S_8	$[(AB)(CD)][(AC)(BD)]$	0	4

We now consider the effect on each of these states of an exchange of partners *occurring between the centromeres and the re-pairing or crossing-over nearest to the centromeres.* Clearly, a re-pairing will turn S_1 into S_1 and a crossing-over (between non-sister strands) will turn S_1 into S_2. However, the outcome is not certain. For example, such a crossing-over may turn S_2 into S_1, S_2, or S_5 and will do so with probabilities $\frac{1}{4}$, $\frac{1}{4}$, and $\frac{1}{2}$ respectively. Writing such an outcome as $\frac{1}{4}S_1 + \frac{1}{4}S_2 + \frac{1}{2}S_5$ and enumerating all the possible cases we obtain Table 2.

TABLE 2.   EFFECT OF RE-PAIRING AND CROSSING-OVER
NEAREST THE CENTROMERE

original state	effect of re-pairing	effect of cross-over
S_1	S_1	S_2
S_2	S_3	$\frac{1}{4}S_1 + \frac{1}{4}S_2 + \frac{1}{2}S_5$
S_3	$\frac{1}{2}S_2 + \frac{1}{2}S_3$	S_4
S_4	S_4	$\frac{1}{4}S_3 + \frac{1}{4}S_4 + \frac{1}{2}S_6$
S_5	S_7	$\frac{1}{2}S_2 + \frac{1}{2}S_5$
S_6	$\frac{1}{2}S_6 + \frac{1}{2}S_8$	$\frac{1}{2}S_4 + \frac{1}{2}S_6$
S_7	$\frac{1}{2}S_5 + \frac{1}{2}S_7$	S_8
S_8	S_6	$\frac{1}{2}S_7 + \frac{1}{2}S_8$

Suppose, now that we have a situation in which re-pairing and crossing-over has occurred in a certain order along the arms of the chromosomes between the locus and the centromere. We can regard the final configuration as seen from the centromeres as the state of a random process of the type of a finite chain. We suppose that we start from a state in which there is no re-pairing and no crossing-over and introduce the latter in order, *beginning with the one nearest the locus*. The process is similar to a Markov chain except that we now have two different types of matrices of transition probabilities corresponding to re-pairing and crossing over.

The state after any number of these can be represented by a column vector

$$\boldsymbol{p} = (p_1, p_2, \ldots, p_8)'$$

whose elements are the probabilities of the states $S_1, \ldots S_8$. The initial state is then

$$\boldsymbol{p_0} = (1, 0, \ldots, 0)',$$

and the final state will be obtained by pre-multiplying this vector by matrices representing the effects of re-pairing and crossing-over, this pre-multiplication being done in the order in which these events occur starting from the locus and moving towards the centromere. Let $\boldsymbol{T_1}$ and $\boldsymbol{T_2}$ denote the transition probability matrices associated with re-pairing and crossing-over respectively. Then from Table 2 we see that

$$\boldsymbol{T_1} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 \end{bmatrix}$$

and

$$\boldsymbol{T_2} = \begin{bmatrix} 0 & \frac{1}{4} & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & \frac{1}{4} & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{4} & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & \frac{1}{4} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & \frac{1}{2} \end{bmatrix}$$

The probabilities of the various possible final states will be then obtained by pre-multiplying $\boldsymbol{p_0}'$ by $\boldsymbol{T_1}$ and $\boldsymbol{T_2}$ in the order in which they occur.

T_1 is a matrix which, by the re-arrangement of rows and columns, can be represented as a diagonal matrix of five matrices, two of these being of the form (1), and three of the form

$$\begin{pmatrix} 0 & \frac{1}{2} \\ 1 & \frac{1}{2} \end{pmatrix}$$

Thus the characteristic roots of T_1 are $\lambda = 1$ (five times) and $\lambda = -\frac{1}{2}$ (three times). Any characteristic post-vector corresponding to the root $\lambda = 1$ will be of the form

$$p = (x_1,\ \tfrac{1}{2}x_3,\ \tfrac{1}{2}x_3,\ x_3,\ \tfrac{1}{2}x_4,\ \tfrac{1}{2}x_5,\ \tfrac{1}{2}x_4,\ \tfrac{1}{2}x_5),'$$

where x_1, x_2, x_3, x_4 and x_5 are arbitrary.

Similarly T_2 can be transformed, by exchange of rows and columns, into a diagonal matrix of three matrices which are

$$\begin{pmatrix} 0 & \frac{1}{2} & 0 \\ 1 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

twice and

$$\begin{pmatrix} 0 & \frac{1}{2} \\ 1 & \frac{1}{2} \end{pmatrix}$$

once. Thus the characteristic roots of T_2 are $\lambda = 1$ (three times), $\lambda = \frac{1}{2}$ (twice), $\lambda = -\frac{1}{2}$ (three times) and any characteristic post-vector for the root $\lambda = 1$ must be of the form

$$p = (\tfrac{1}{2}y_1,\ \tfrac{1}{2}y_1,\ \tfrac{1}{2}y_2,\ \tfrac{1}{2}y_2,\ \tfrac{1}{2}y_1,\ \tfrac{1}{2}y_2,\ \tfrac{1}{2}y_3,\ \tfrac{1}{2}y_3).'$$

If we compare this with the general form of the characteristic post-vector for $\lambda = 1$ of T_1 we see that any characteristic post-vector ($\lambda = 1$) of both matrices must be a multiple of

$$(1,\ 4,\ 8,\ 32,\ 4,\ 32,\ 8,\ 16)'$$

and since to represent probabilities the sum of the elements must add to unity, must therefore be

$$\pi = \left(\frac{1}{105},\ \frac{4}{105},\ \frac{8}{105},\ \frac{32}{105},\ \frac{4}{105},\ \frac{32}{105},\ \frac{8}{105},\ \frac{16}{105} \right).'$$

We now enquire under what circumstances an unlimited repetition of the operations T_1 and T_2 result in a convergence of the vector representing the probabilities to π.

That this cannot hold in general can be seen as follows. Consider first the operation $T_1 T_2$ performed indefinitely often on the initial vector p_0. Since π is a

characteristic post-vector of both T_1 and T_2 for $\lambda = 1$ it is also a characteristic post-vector for $T_1 T_2$. We can show that it is the sole post-vector for $\lambda = 1$ so that unity is a simple root of $T_1 T_2$ by considering the accessibility of states in each operation. Thus in the operation T_2 state 2 is accessible from state 1 and is the only such state whilst states 1, 2, and 5 are accessible from state 2 and so on. Table 3 shows the states accesible from each state in the operations T_1 and T_2 and also in the operation $T_1 T_2$.

TABLE 3. ACCESSIBLE STATES

initial state	state accessible with		
	T_1	T_2	$T_1 T_2$
1	1	2	3
2	3	1, 2, 5	1, 3, 7
3	2, 3	4	4
4	4	3, 4, 6	2, 3, 4, 6, 8
5	7	2, 5	3, 7
6	6, 8	4, 6	4, 6, 8
7	5, 7	8	6
8	6	7, 8	5, 6, 7

From this it can be verified that $(T_1 T_2)^4$ is such that every state is accessible from every state. Such a matrix is said to be positively regular and is known to have only a single root of unit modulus. Thus

$$(T_1 T_2)^\infty \, p_0 = \lim_{n \to \infty} (T_1 T_2)^n \, p_0 = \pi$$

and similarly $(T_2 T_1)^4$ has all its elements non-zero and is thus also positively regular. On the other hand it is easily seen that

$$\lim_{m \to \infty} \lim_{n \to \infty} T_1^m \, T_2^n \, p_0 \neq \pi$$

since it can be verified that for no values of m and n are states 4, 6 and 8 accessible by means of $T_1^m \, T_2^n$ from state 1. In a similar way

$$\lim_{n \to \infty} \lim_{m \to \infty} T_2^n \, T_1^m \, p_0 \neq \pi.$$

Thus we have a situation in which the ordinary limiting behaviour of a Markov chain, with a single matrix of transition probabilities, does not occur.

Only with some further restrictions on the order of application of T_1 and T_2 can we conclude that after a number of repetitions of T_1 and T_2 will the probabilities of the various states be given by π. We have already seen that this will occur if they

alternate. This is not the sort of condition which is plausible biologically. Suppose that we assume that n exchanges of partner occur between the centromere and the locus and that in the $n+1$ intervals formed by these crossing-over occurs with frequencies which have the same discrete non-degenerate probability distribution. Then clearly by the same type of argument as used above, as n increases, the frequency distribution of crossing-over remaining unaltered, the final distribution will tend to π. Thus when we get well away from the centromere the probabilities of the various states will be well approximated by the elements of π.

We shall now show that if this is so, segregation is of the chromatid type and also see how the parameter α can be expressed in terms of the sequence of T_i's. We have to calculate the probabilities $p(AA),\dots$ of obtaining gametes of the form $AA,\dots,$ $DD, AB, AC, \dots CD$. From the initial assumption that each state S_i corresponds to all possible permutations of the symbols $A, B, C,$ and D with equal probabilities we have

$$p(AA) = p(BB) = p(CC) = p(DD) = \tfrac{1}{4}\alpha, \text{ say,}$$

and $$p(AC) = \dots = p(CD) = \tfrac{1}{4}(1-\alpha).$$

Each state will have a different value of α which we write as α_i. By straightforward examination we can evaluate these as in Table 4.

TABLE 4. VALUES OF α

state	α	state	α
S_1	0	S_5	$\tfrac{1}{2}$
S_2	$\tfrac{1}{12}$	S_6	$\tfrac{1}{3}$
S_3	$\tfrac{1}{12}$	S_7	$\tfrac{2}{3}$
S_4	$\tfrac{2}{3}$	S_8	$\tfrac{1}{2}$

If we write these values as a row vector

$$\alpha' = (0, \tfrac{1}{12}, \tfrac{1}{12}, \tfrac{2}{3}, \tfrac{1}{2}, \tfrac{1}{3}, \tfrac{2}{3}, \tfrac{1}{2})$$

and the probabilities of the states are given by a column vector p, the value of α will be

$$\alpha = \alpha' p.$$

In particular we see that $$\alpha'\pi = \tfrac{1}{7}$$

and this corresponds to chromatid segregation for which $p(AA) = \dots = p(DD) = (\tfrac{8}{2})^{-1} = \tfrac{1}{28}$. Thus in this way we can find the correct value of α corresponding to any specified sequence of re-pairing and crossing-over.

POLYSOMIC INHERITANCE AND THE THEORY OF SHUFFLING

SOME GENERAL PROBLEMS

The above results suggest some general problems about what may be called non-homogeneous processes of Markov chain type. In the problem considered above we had two matrices of transition probabilities both of which had multiple unit roots, and a single common right characteristic vector for the value $\lambda = 1$. Consider a simpler situation in which we have a process which can be in any of n states and two matrices of transition probabilities, $T_1 = (p_{ij})$ and $T_2 = (p_{ij}')$ which each have only a single simple root of modulus unity (which must equal unity), and a common right characteristic vector. Assume further, for the sake of simplicity, that for each of these matrices the other roots are unequal. Let $\lambda = 1, \lambda_2, ..., \lambda_n$ be the roots of T_1, with right characteristic vectors $t_1, ... t_n$ and similarly $\mu_1 = 1, \mu_2, ..., \mu_n$ and $u_1 = t_1$, $u_2, ... u_n$ for T_2.

Then the t_i are linearly independent and if p_0 is any initial vector of probabilities we can write

$$p_0 = t_1 + \sum_{i=2}^{n} \alpha_i t_i$$

so that

$$T_1^N p_0 = t_1 + \sum_{i=2}^{n} \alpha_i \lambda_i^N t_i.$$

Then as N tends to infinity, $T_1^N p_0$ converges to t_1. In fact since we can write the difference between these as

$$T_1^N p_0 - t_1 = \sum_{i=2}^{n} \alpha_i \lambda_i^N t_i$$

we see that, in a certain sense, $T_1 p$ is closer to t_1 than p, for any vector p of probabilities. The sense in which this is true is that the length of the vector $p - t_1$ is always reduced by the operator T_1, when the length is defined in terms of the coordinate system defined by the n vectors $t_1, ... t_n$ which span the whole space, since they are linearly independent.

A similar result holds for the operator T_2 which reduces in length any vector $p - t_1 = p - u_1$, p a vector of probabilities, the length being now defined in terms of the coordinate system defined by the n linearly independent vectors $u_1, ... u_n$.

These results suggest the following conjecture. Suppose that p_0 is any vector of probabilities and that

$$S_N = T_{i_N} T_{i_{N-1}} \cdots T_{i_2} T_{i_1}$$

where $i_N, ... i_1$ is any fixed sequence of 1's and 2's.

Conjecture. As N tends to infinity, $S_N p_0$ tends to t_1 for any prescribed sequence $i_1, ... i_N, ...$

This result is false. Consider a process with four states, and two transition matrices

$$T_1 = \begin{bmatrix} 0 & 0 & \tfrac{1}{2} & \tfrac{1}{2} \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \tfrac{1}{2} & \tfrac{1}{2} \end{bmatrix}, \quad T_2 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & \tfrac{1}{2} & \tfrac{1}{2} \\ 1 & 0 & 0 & 0 \\ 0 & 0 & \tfrac{1}{2} & \tfrac{1}{2} \end{bmatrix}$$

T_2 is obtained from T_1 by interchanging the first two rows and the first two columns. The characteristic equation of T_1 is

$$\lambda(\lambda-1)(\lambda^2+\tfrac{1}{2}\lambda+\tfrac{1}{2}) = 0$$

and the roots are $\lambda = 1$, $\lambda = -\tfrac{1}{4}(1 \pm \sqrt{-3})$, $\lambda = 0$ so that there is only one root of modulus unity. The characteristic equation of T_2 is the same and both have the vector $(\tfrac{1}{4}, \tfrac{1}{4}, \tfrac{1}{4}, \tfrac{1}{4})'$ as the right characteristic vector corresponding to $\lambda = 1$, as is otherwise obvious since both matrices are doubly stochastic.

Then

$$T_1 T_2 = \begin{bmatrix} \tfrac{1}{2} & 0 & \tfrac{1}{4} & \tfrac{1}{4} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \tfrac{1}{2} & \tfrac{1}{2} \\ \tfrac{1}{2} & 0 & \tfrac{1}{4} & \tfrac{1}{4} \end{bmatrix}$$

This has the characteristic equation

$$\lambda(\lambda-1)^2(\lambda-\tfrac{1}{4}) = 0$$

so that $\lambda = 1$ is a double root with two right characteristic vectors which can be taken as $(\tfrac{1}{3}, 0, \tfrac{1}{3}, \tfrac{1}{3})$ and $(0, 1, 0, 0)$. Then $(T_1 T_2)^n p_0$ will have a limiting behaviour which depends on p_0.

It would, therefore, be of great interest to investigate what conditions imposed on either the order of the operations, or on the matrices themselves, will result in convergence to a unique vector independent of the initial conditions. In particular, this would have useful application to the case where the states are the $n!$ permutations of n objects and the operations are probability mixtures of the permutation operations.

This work was partly carried out whilst the author was at Princeton University under contract with the Office of Naval Research.

REFERENCE

FISHER, R. A. and MATHER, K. (1943): The inheritance of style length in Lythrum salicaria. *Ann. Eugen.*, **12**, 1-23.

Paper received : August, 1960.

MISCELLANEOUS

APPARENT ANOMALIES AND IRREGULARITIES IN MAXIMUM LIKELIHOOD ESTIMATION*

(with discussion)

By C. RADHAKRISHNA RAO
Indian Statistical Institute, Calcutta

1. INTRODUCTION

Maximum likelihood (m.l.) estimation is criticised mainly on the following grounds:

(i) It does not always provide consistent estimates.

(ii) There exist estimates with lower asymptotic variance than that of the m.l. estimate, and therefore the m.l. method does not lead to most efficient estimates as claimed in the literature on this subject.

(iii) The computations involved in determining m.l. estimates are in most cases unduly heavy. On the other hand, there exist simpler methods of estimation which provide estimates which are asymptotically as efficient as m.l. estimates.

(iv) No adequate justification has been put forward for m.l. estimation in finite samples. Judged by the criterion of mean squared error in finite samples, there are examples where certain other procedures are better than m.l.

The criticism (iii) on grounds of computational difficulties will be relatively unimportant when high speed electronic computers become easily available for use by research workers. The computations, involving an iterative procedure for solving m.l. equations and inversion of matrices for obtaining standard errors, can be easily programmed on any modern electronic computer. Recently, routine programmes have been constructed at the Indian Statistical Institute for obtaining m.l. estimates of gene frequencies, standard errors, expected frequencies and goodness of fit χ^2, from observed phenotypic frequencies of various blood group systems such as OAB, MN, CDE, etc. The time taken for these computations is of the order of a minute for each blood group system, even on a comparatively slow machine like the HEC (Hollerith Electronic Computer).

I shall, therefore, confine my comments to the other points of criticism relating to consistency, efficiency; and properties of estimates in small samples.

2. PURPOSE OF ESTIMATION

It will help in our discussion if we agree on the purpose of estimation, on which will depend the criteria for the choice of a suitable method of estimation. Much of the controversy in the literature on estimation could be dismissed once this problem is properly answered. There has been a tendency to consider estimation as a part of decision theory, which requires as a datum of the problem the specification of the loss for a given difference between the estimate and the true value of the unknown parameter. The criterion in such a case is naturally the minimisation of expected loss. This may be appropriate in certain situations but I

---

* This paper originally presented at the 32nd Session of the International Statistical Institute is being reprinted here with the kind permission of the Editor, *Bulletin of the International Statistical Institute.*

am not sure whether one can support Berkson (1955) when he wants to estimate the slope of the probit regression line in a bio-assay using the criterion of minimum expected squared error, unless of course he believes or makes us believe that the loss to society is proportional to the square of the error in his estimate. I suppose a bio-assayer, when he obtains an esti- mate of the standard deviation of a tolerance distribution, or the LD 50, uses it in a variety of ways besides playing a game with nature or with society.[1] He would like to compare it with an estimate of LD 50 for another insecticide, combine it with a previous estimate for the same insecticide to obtain a better estimate, preserve it for comparison or combination with future estimates, or indulge in some assertions (with some confidence) that LD 50 is less than a specified value or lies between two specified values and so on, or use the estimate itself more conveniently in the place of basic data in reaching optimum decisions for a specified loss function.

It may be argued that all these problems could be answered directly, and in theory more satisfactorily, from given data without considering the intermediate methodological problem of estimation. If then we insist on estimating the unknown parameters and use the estimates for purposes of inference such as those indicated above it can only be due to some convenience in handling the estimates rather than the original data, in addition to the resulting economy in recording only the estimates for future use, instead of preserving the entire mass of observed data, much of which may be irrelevant. If, therefore, we define the purpose of estimation as condensation of data, what criteria can we lay down for choosing a method of estimation ?

Most statisticians would probably agree that statistical inference consists, in general, in discriminating between alternative possible situations on the basis of given data, and as such it should be based on the likelihood $P(S, \theta)$ of the parameter θ given the sample S, which is same as the probability (or probability density) of S given θ. More precisely, we need the ratio of the likelihoods for two given values θ_1 and θ_2 of the parameter. There may be, however, some controversy about the form in which the uncertainty in the choice of θ_1 or θ_2 given S, is to be expressed.

If there exists a statistic T such that

$$P(S, \theta_1)/P(S, \theta_2) = P(T, \theta_1)/P(T, \theta_2) \qquad \qquad \ldots \ (2.1)$$

for all admissible θ_1 and θ_2, nothing is lost by replacing the sample S by the statistic T, which is for all relevant purposes equivalent to S. Such a statistic T is said to be sufficient in the sense of Fisher (1922). There will be a multiplicity of statistics T satisfying (2.1), one of them (in the extended sense of the term statistic) being the sample itself. In general we can choose one among them, say T_0, which is minimal in the sense that T_0 is essentially a function of every sufficient statistic T (Lehmann and Scheffé, 1950). A minimal sufficient statistic thus provides an exhaustive summary of the sample for purposes of statistical in- ference. If $x_1, ..., x_n$ is a sample of observations, the observed mean $\bar{x}$ and variance s^2 are jointly sufficient when the population distribution is normal with unknown mean and variance.

---

[1]A simple example given by Silverstone (1957) illustrates the point. In the case of the ordinary binomial distribution with probability θ, for number of trials $n = 3$, the estimate $T_1 \equiv 1 \ 2$ of θ has smaller mean square error than the observed proportion $r \ n$ for all true values of θ between $1 \ 4$ and $3 \ 4$. This cannot be advanced as a cogent reason for using T_1 instead of r/n when nothing is known about the true value of θ.

On the other hand, the minimal sufficient statistic may be the ordered values in the whole sample as in the case of a Cauchy population with an unknown location parameter, i.e. no reduction of the data is possible without disturbing the relation (2.1). In such a case we may look for a statistic which belongs to a specified simple class and provides the maximum possible discrimination.

For a statistic T belonging to a specified class let us denote the likelihood ratio by $P(T, \theta_1)/P(T, \theta_2)$. The larger the deviation of this ratio from 1, the greater will be the discrimination between the parameters θ_1 and θ_2. For purposes of comparison, it is convenient to have a measure of the amount of discrimination provided by a statistic T. There is, indeed, some amount of arbitrariness in the choice of such a measure. Some natural measures are, 'the amount of overlap' between distributions corresponding to θ_1 and θ_2 as defined by the author (Rao, 1948), or the quantity

$$J_T(\theta_1, \theta_2) = \operatorname*{E}_{\theta_1} \log \left\{ \frac{P(T, \theta_1)}{P(T, \theta_2)} \right\} + \operatorname*{E}_{\theta_2} \log \left\{ \frac{P(T, \theta_2)}{P(T, \theta_1)} \right\} \qquad \ldots \ (2.2)$$

considered by Kullback and Liebler (1951) following the concepts of information theory. One may prefer even a pure distance measure like

$$\rho_T(\theta_1, \theta_2) = \int \sqrt{P(T, \theta_1) \, P(T, \theta_2)} \, dT \qquad \ldots \ (2.3)$$

introduced by Hellinger (1909) (see also Bhattacharya, 1946). Each of these measures is not more than the corresponding expression when T is replaced by the whole sample. The ratio of the amount of discrimination provided by T to that contained in the whole sample may be considered as an index of the effectiveness of T. When T is sufficient this ratio is unity for all these measures. For simplicity, let us consider $J_T(\theta_1, \theta_2)$ in the further discussion, observing that the same or similar results will be valid for the other measures mentioned.

When the sample S consists of n independent observations on a variate X we have

$$J_S(\theta_1, \theta_2) = n \left\{ \operatorname*{E}_{\theta_1} \log \left[P(X, \theta_1)/P(X, \theta_2) \right] + \operatorname*{E}_{\theta_2} \log \left[P(X, \theta_2)/P(X, \theta_1) \right] \right\} \qquad \ldots \ (2.4)$$

where the expression within the brackets is the value of $J(\theta_1, \theta_2)$ for a single observation and is therefore independent of n. As $n \to \infty$, $J_S(\theta_1, \theta_2) \to \infty$, and we have perfect discrimination between θ_1 and θ_2, as was shown by Basu (1954). A rigorous demonstration of this result was given earlier by Kakutani (1948) using the fact that $\rho_S(\theta_1, \theta_2) \to 0$ as $n \to \infty$. He showed that the distributions of the sample sequences in the infinite dimensional space for two different values of θ are 'orthogonal'. A statistic T_n which replaces a sample would not be of much use if it did not provide complete discrimination between any two values of θ as $n \to \infty$, i.e., if $P(T_n, \theta_1)$ and $P(T_n, \theta_2)$ are not orthogonal in the limit. This is possible if[2] $T_n \to \phi(\theta)$ with probability 1 as $n \to \infty$, where $\phi(\theta)$ is a function of θ, having one-to-one correspondence with the possible values of θ. This is exactly what the criterion of consistency[3] laid down by Fisher (1922) demands.

---

[2] Generally, orthogonality is possible only if T_n tends to a particular value but examples may be found where for each θ, T_n has a non-degenerate limiting distribution, with distributions corresponding to different values of θ being non-overlapping.

[3] We are not demanding that $T_n \to \theta$. It is enough, if for any two different values of θ, T_n tends to two different constants. In such a case T_n is defined to be consistent for θ in the wide sense (see section 4 of this paper).

For given n, $J_S(\theta_1, \theta_2)$ and $J_T(\theta_1, \theta_2)$ depend on how widely separated are the distributions corresponding to θ_1 and θ_2. Therefore, the ratio of $J_T(\theta_1, \theta_2)$ to $J_S(\theta_1, \theta_2)$ may not represent the true effect of replacing S by T if the distributions corresponding to θ_1 and θ_2 are widely different. We may, therefore, consider the ratio of these quantities as $\theta_2 \to \theta_1$ assuming that this implies closeness of distributions. It is easy to see that

$$J_S(\theta_1, \theta_1 + \delta\theta_1) \sim \frac{n}{2}\, i(\theta_1)(\delta\theta_1)^2$$

$$J_T(\theta_1, \theta_2 + \delta\theta_1) \sim \frac{n}{2}\, i_T(\theta_1)(\delta\theta_1)^2 \qquad \dots \ (2.5)$$

where $i(\theta_1) = E_{\theta_1}[P'(x, \theta_1)/P(x, \theta_1)]^2$ is the information per observation as defined by Fisher (1922, 1925) and $i_T(\theta_1)$ the corresponding information per observation in the statistic T. It is shown by Fisher (1925) that $i_T(\theta_1) \leqslant i(\theta_1)$, which suggests the criterion of maximising the information per observation in the choice of a statistic.

Reference may also be made to earlier work by the author (Rao, 1945) where the distance between two distributions differing by small quantities in the parameters is derived, by an argument similar to that used here, as a quadratic differential metric of which (2.5) is a special case. In the general case $i(\theta_1)$ and $i_T(\theta_1)$ are matrices, and it is known that $\{i(\theta_1) - i_T(\theta_1)\}$ is a positive semi-definite matrix. The efficiency of a statistic may be measured by some expression reflecting the deviations from zero of the elements in the matrix $\{i(\theta_1) - i_T(\theta_1)\}$. We shall consider only the single parameter case in further discussions.

The information function seldom provides a complete ordering of the statistics for all values of θ in the admissible range. It is, of course, possible to obtain a complete ordering with respect to the average amount of information based on an *a priori* distribution of θ, if this last distribution can be specified. In other situations no satisfactory solution seems to exist, although information can be used to eliminate some statitstics which are worse than others in a range of the parameters in which we are interested. Fortunately under favourable circumstances, there exist methods of estimation for which $i_T(\theta) \to i(\theta)$ as $n \to \infty$, so that we have an assurance that at least in large samples the relative information lost is small.

In small samples, we could examine the performance of any statistic by computing the ratio i_T/i. If this quantity is small, we need not insist on replacing the observations S by the statistic T, but strengthen T by considering other statistics in addition to T, so that all taken together provide information per observation comparable to i. In the worst case, when the sample size is small, it may be necessary to retain the entire sample or the likelihood function either in the form of a graph or tabulated for some values of the parameters, which would enable us to reconstruct the function without much error, if needed in future.

3. Efficiency

3.1. *A new formulation of the concept of efficiency.*

Having discussed certain broad principles for summarising data we may examine some easily recognisable properties of statistics by which we can judge their effectiveness and discuss methods by which such statistics are obtained.

APPARENT ANOMALIES AND IRREGULARITIES IN M. L. ESTIMATION

Let us consider the consequences of replacing the observations by a statistic in discriminating an alternative value of a parameter close to a specified value θ. It was shown by Rao and Poti (1946) by an application of the important lemma of Neyman and Pearson that a test which discriminates best small departures from a given value of θ, for *any given* n, is of the form : "Reject if and only if $Z_n \geqslant \lambda$" where

$$Z_n = (\sqrt{n})^{-1} \sum_1^n z_t, \quad z_t = P'(x_t, \theta)/P(x_t, \theta) \qquad \ldots \ (3.11)$$

λ is a constant, $P(x, \theta)$ is the probability density of x, and $x_1, \ldots, x_n$ are n independent observations. Or, if we denote by L the likelihood of the parameter given the sample, the statistic Z_n is simply $(d \log L/d\theta)/\sqrt{n}$.

Can we construct a statistic independent of θ and with a performance[4] as good as that of Z_n? This is possible when there exists a function T_n of the observations such that

$$Z_n = \lambda(\theta) f(T_n) + \mu(\theta) \qquad \ldots \ (3.12)$$

or, more generally, when the variance of Z_n given T_n is zero, a situation in which T_n is sufficient for θ. On the other hand, it may be possible to construct a statistic such that its asymptotic correlation with Z_n is unity as $n \rightarrow \infty$. Such a statistic, if it exists[5] is as good as Z_n in sufficiently large samples, i.e., is best for discrimination between two neighbouring values of the parameter in sufficiently large samples. Based on these considerations we give a new formation of the concept of efficiency.

Definitions. A statistic is said to be efficient if its asymptotic correlation with the derivative of log likelihood is unity. The efficiency of any statistic may be measured by ρ^2, where ρ is its asymptotic correlation with Z_n.

In the case of more than one unknown parameter, a statistic consistent for a parameter is said to be efficient if its multiple correlation with the derivatives of the log likelihood with respect to the unknown parameters is unity. The efficiency of any statistic is measured by the square of the multiple correlation.

3.2. *'Super efficient' estimates and their efficiency.*

An efficient statistic is defined by Fisher (1922) as one whose asymptotic variance is $[n \, i(\theta)]^{-1}$, or alternatively as one whose asymptotic variance is the least. Although Fisher formally stated the criterion of efficiency in terms of least asymptotic variance it is clear from his writings that by an efficient estimate he meant a statistic for which the loss of information per observation tends to zero. Fisher gives the following extended definition of efficiency on page 714 of his 1925 paper : 'The efficiency of a statistic is the ratio of the intrinsic accuracy of its random sampling distribution to the amount of information in the data from which it has been derived.' He argued that since the reciprocal of information for the mean of the

---

[4] The power of the test based on Z_n, when n is large and the alternative to θ is $\theta + \delta\theta$, is nearly $\phi[n \, i(\theta)]d\theta$ where ϕ is an increasing function of the argument. The quantity $i(\theta)$ which appears in the expression (2.5) for the distance between distributions close to one another is also explicitly involved in the power function.

[5] For instance the unique consistent root T_n^* of the m.l. equation (ref. Huzurbazar, 1948) under the conditions given by Doob (1934) or Cramer (1946) satisfies that property, for $|\sqrt{n}(T_n - \theta) - Z_n| \rightarrow 0$ with probability 1. An m.l. estimate when referred to in the sequel is assumed to have this property.

distribution is variance when the distribution is normal and the information in a statistic is bounded above by $n\, i(\theta)$, an efficient statistic is recognised when it has a limiting normal distribution with variance $[n\, i(\theta)]^{-1}$ which is the least possible. In 1951, J. L. Hodges[6] and later Le Cam (1953) constructed examples of consistent estimates with an asymptotic variance $\leqslant [n\, i(\theta)]^{-1}$, with strict inequality for certain values of θ. In fact, at these exceptional points the asymptotic variance can be made arbitrarily small. These examples of what are called 'super efficient' estimates show that there is no non-zero bound to the asymptotic variance of a consistent estimate, contrary to what is stated by Fisher. One might think that super efficient estimates with asymptotic variance $\leqslant [n\, i(\theta)]^{-1}$ should be preferred to efficient estimates with asymptotic variance $[n\, i(\theta)]^{-1}$. We shall examine these notions in the light of the new definition of efficiency given here.

First it may be noted that super efficiency arises, when the statistic is not an explicit function of the sample distribution function and therefore not satisfying the consistency condition as originally defined by Fisher.[7] Assuming Fisher consistency (FC) and certain regularity conditions (mainly Frechet differentiability) on the statistic, Kallianpur and Rao (1955) demonstrated that $[n\, i(\theta)]^{-1}$ is, indeed, a lower bound to the asymptotic variance, thus justifying Fisher's argument. It is also deducible from the results of Kallianpur and Rao that a FC statistic with asymptotic variance $[n\, i(\theta)]^{-1}$, under the regularity conditions assumed, has asymptotic correlation unity with Z_n. This demonstrates the equivalence of Fisher's definition of efficiency with that proposed here under the regularity conditions imposed on the estimate. Earlier work by Neyman (1949) and Barankin and Gurland (1950) also tend to confirm Fisher's results.

Now let us see how the new definition of efficiency enables us to judge the effectiveness of any statistic, whether it satisfies regularity conditions or not. What happens when FC and other regularity conditions imposed on the statistic are not satisfied ? In this case 'super efficient' estimates do exist as shown by Hodges and Le Cam. We shall show that when a super efficient estimate (i.e. with a possibly smaller asymptotic variance than that of the m.l. estimate) exists, one of the following two possibilities holds.

---

[6] The example by Hodges is quoted in a paper by Le Cam (1953). Consider the mean X_n of n independent observations on X from a normal distribution with mean θ and standard deviation unity. As is well known X_n is the m.l. estimate of the mean with variance $\sigma_n^2 = 1/n$. Let T_n be the function defined by

$$T_n(X_n) = X_n \quad \text{if } |X_n| > \frac{1}{n^{1/4}}$$

$$= \alpha X_n \text{ if } |X_n| < \frac{1}{n^{1/4}}$$

It is easy to see that T_n is also asymptotically normally distributed about θ, with variance $= 1/n$ for $\theta \neq 0$, and α^2/n for $\theta = 0$. Since α is arbitrary, the asymptotic variance is less than that of the m.l. estimate when $\theta = 0$. This example of Hodges was generalized by Le Cam to improve the asymptotic variance at a countable number of values of the parameter θ.

[7] If S_n is the sample distribution function and $F(\theta)$ the true distribution function a functional $f(S_n)$ is said to be Fisher consistent (FC) for θ if $f\{F(\theta)\} \equiv 0$. For a discussion on this subject see Kallianpur and Rao (1955). The estimates of Hodges and Le Cam are not FC.

(1) In large samples, it is equivalent to the m.l. estimate (which is efficient in the new sense), i.e., has asymptotic correlation unity with the m.l. estimate, and therefore it is efficient in the new sense.

(2) It is not efficient in the new sense, in which case it is definitely worse than the m.l. estimate, for purposes of inference such as testing of hypothesis, interval estimation, etc.

Let us assume that a statistic T_n consistent for θ is such that $\sqrt{n}(T_n - \theta)$ and Z_n have a joint asymptotic distribution. Denote the asymptotic variance of $\sqrt{n}(T_n - \theta)$ by $v(T)$ and its asymptotic covariance with Z_n by $\alpha(\theta)$. Since the asymptotic variance of Z_n is $i(\theta)$ we have the obvious inequality

$$v(T) \geqslant \alpha^2(\theta)/i(\theta), \qquad \qquad \ldots \ (3.21)$$

From this relation, it follows that T_n has asymptotic correlation unity with Z_n, or it is fully efficient (in the new sense) if and only if $v(T) = \alpha^2(\theta)/i(\theta)$. If T_n^* is an m.l. estimate for which the observation made in footnote (5) is true, then $\sqrt{n}(T_n^* - \theta)$ has asymptotic variance equal to $1/i(\theta)$, and asymptotic correlation unity with Z_n. Therefore, when the equality in (3.21) is attained, T and T^* have asymptotic correlation unity *whatever* may be the inequality satisfied between their asymptotic variances[8], $\alpha^2(\theta)/i(\theta)$ and $1/i(\theta)$. If $\alpha^2(\theta) \leqslant 1$ for all θ, with strict inequality for some θ, we have an example of super efficiency as in the case of Hodges' example.[9] In fact, we can use the device of Hodges to construct examples of 'sub efficiency' i.e. where $\alpha^2(\theta) \geqslant 1$. In either case, when the equality in (3.21) is attained, T_n' is equivalent to the m.l. estimate T_n^* in the sense that essentially the same type of inference is possible by using T_n or T_n^* in large samples, whether T_n is super or sub efficient in the earlier sense.

We may now ask what happens when the equality in (3.21) is not attained. Such a statistic T_n has asymptotic correlation $-1 < \rho < 1$ with Z_n, and therefore is not as good as Z_n (and therefore not as good as m.l.) for local discrimination, although T_n may be super efficient, i.e.

$$\frac{1}{i(\theta)} \geqslant v(T) > \frac{\alpha^2(\theta)}{i(\theta)}. \qquad \qquad \ldots \ (3.22)$$

Consider for example a sample $x_1, \ldots, x_n$ from a normal distribution with an unknown mean μ and variance unity and denote by $\bar{x}$ and x_m, the sample mean and median respectively. Define the statistic.

$$\begin{aligned} T &= \alpha x_m & \text{if } \bar{x} < n^{-1/4} \\ &= \bar{x} & \text{if } \bar{x} \geqslant n^{-1/4} \end{aligned} \qquad \ldots \ (3.23)$$

It is easy to see that the asymptotic distribution of T is normal with variance $\alpha^2 \pi/2$ when $\mu = 0$, and 1 when $\mu \neq 0$. By choosing α arbitrarily small, $\alpha^2 \pi/2$ can be made less than 1. The statistic T is therefore super efficient. But for testing the hypothesis $\mu = 0$, it is obvious that the test criterion is essentially the median when the null hypothesis is true and consequently the power of the test is smaller than that of $\bar{x}$. In this connection we may also refer

---

[8] We may compare this result with that of Fisher (1925), that the asymptotic correlation between two efficient estimates having the same least asymptotic variance is unity.

[9] It may be observed from the example given in footnote (6) that T_n and X_n have asymptotic covariance α for $\theta = 0$ and 1 for $\theta \neq 0$. The value of α, can be chosen to be > 1 or < 1 arbitrarily. The technique of Hodges and Le Cam provides a statistic which is essentially equivalent to the statistic with which they start.

to an interesting but a different type of example due to Basu (1956), where the ratio of a limiting variance of one statistic to that of another $\to \infty$ but the corresponding ratio of the probabilities of concentration wthin any given limits of the true value $\to 0$. So it would appear that the criterion of minimum asymptotic variance is misleading.

We may now raise the problem whether given a super efficient estimate, it is possible to find a function of the m.l. estimate which is consistent for the parameter and which has a smaller asymptotic variance than the given super efficient estimate. This means that a super efficient estimate can be uniformly improved from the point of view of asymptotic variance by using a function of the m.l. estimate. This is true of the known examples of super efficiency. Further, given a super efficient estimate, i.e., when $v(T)$ satisfies (3.22), we can construct a function of m.l. estimate, by using Le Cam's technique, such that its asymptotic variance is smaller than $v(T)$ or even $\alpha^2(\theta)/i$ at a countable set of values of θ. To examine whether improvement is possible for all values of θ, we have to study the function $x(\theta)$. Under some assumptions Le Cam (1953) proved that $|x(\theta)|$ can be less than unity only for a set of points of Lebesgue measure zero. This is encouraging but does not solve the problem posed here. We may have to explore the asymptotic sufficiency of the m.l. estimate (Wald, 1943; Le Cam, 1953) to prove this property.

3.3. *Information in the limit.*

We shall examine the limiting information contained in an efficient estimate, i.e., one which has asymptotic correlation unity with the first derivative of the log likelihood. Let T_n be such an estimate whether it is super or sub efficient with respect to the asymptotic variance. Suppose further that

$$(\sqrt{n}(T_n-\theta), Z_n) \to (T, Z) \text{ in distribution} \qquad \ldots \text{ (3.31)}$$

where (T, Z) is bivariate normal with mean zero and covariance matrix

$$\begin{bmatrix} \sigma_T^2 & \sigma_T\sqrt{i} \\ \sigma_T\sqrt{i} & i \end{bmatrix} \qquad \ldots \text{ (3.32)}$$

Then the variable $\left[T - \dfrac{\sigma_T}{\sqrt{i}} Z\right]$ has zero variance. Therefore

$$\left\{ \sqrt{n}(T_n-\theta) - \frac{\sigma_T}{\sqrt{i}} Z_n \right\} \to 0 \text{ in probability.} \qquad \ldots \text{ (3.33)}$$

For a statistic T_n which satisfies the condition (3.33), under some regularity conditions on $P(x, \theta)$, the probability (or density) of a single observation, Doob (1936) has demonstrated[10] that

$$\lim_{n \to \infty} \{i_{T_n}(\theta)\} = i(\theta) \qquad \ldots \text{ (3.34)}$$

where $i_{T_n}(\theta)$ is the information per observation contained in the statistic T_n computed in the usual way. A simple proof of Doob's proposition is given in a recent paper by the author (Rao, 1960).

---

[10] Doob (1936) states the required condition in terms of strong convergence. I believe this is not necessary.

We have no such assurance about the limiting information in the case of estimates not efficient in the new sense. In fact, if the asymptotic correlation between T_n and Z_n is ρ and that between T_n and $P'(T_n, \theta)/P(T_n, \theta)$ is nearly unity (as may be expected) we have the relation

$$\lim_{n \to \infty} (i_{T_n}) = \rho^2 i \qquad \qquad \dots (3.35)$$

emphasizing the importance of ρ^2 as a measure of efficiency mentioned in the definition of Section 3.1. The use of T_n entails a loss of information equal to that contained in a fraction $(1 - \rho^2)$ of the observations.

3.4. *Efficiency in non-regular cases.*

In non-regular cases such as the rectangular distribution over the range $(0, \theta)$, i.e., where the probability measures corresponding to different values of the parameters are not equivalent, the quantity $i(\theta)$ is not properly defined so that the foregoing theory is not applicable. We shall not discuss such situations in full generality but only consider a special example given by Basu (1952), where the maximum likelihood estimate has a uniformly larger variance than an alternative estimate proposed by him.

Let $x_1, \dots, x_n$ be n observations from a rectangular distribution in the range $(\theta, 2\theta)$, where $0 < \theta < \infty$. The maximum y and the minimum z of the observations are jointly sufficient for θ and the m.l. estimate of θ is $T_1 = y/2$. The asymptotic variance of T_1 is $1/4n^2$ while that of $T_2 = (2y+z)/5$, which is also consistent for θ, is $1/5n^2$. Judged by the criterion of ratio of asymptotic variances the m.l. estimate has only 80% efficiency compared to the alternative estimate. One might be tempted to infer that discrimination based on T_2 is therefore better than that based on T_1, the m.l. estimate, for small differences in the parameter. A computation of the power functions of the tests based on T_1 and T_2 for any sample size shows however that for alternatives close to a given value of θ the power of T_1 is much higher than that of T_2, although T_2 has smaller asymptotic variance than T_1. On the other hand, for alternatives not close to a given value of θ, T_2 is better than T_1.

3.5. *Concentration.*

As is stated above, in the absence of regularity conditions on an estimate T_n, the asymptotic or actual variance of T_n does not necessarily give a good indication of the concentraton of T_n about the true value of θ. An approach to estimation which is concerned explicitly with comparing concentrations has been given recently by Bahadur (1960). This approach may be outlined as follows. Let T_n be a consistent and asymptotically normal estimate of θ based on n independent and identically distributed observations. For any n and any $\epsilon > 0$, let $\tau = \tau(T_n, \epsilon, \theta)$ be defined by the equation

$$P(|T_n - \theta| \geqslant \epsilon | \theta) = 2 \int_{\epsilon/\tau}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt. \qquad \qquad \dots (3.51)$$

τ is called the 'effective standard deviation' of T_n, when θ obtains. It is shown by Bahadur, under mild regularity conditions on the sample space of single observation that

$$\varliminf_{\epsilon \to 0} \lim_{n \to \infty} \{n\, \tau^2\} \geqslant \frac{1}{i(\theta)} \qquad \qquad \dots (3.52)$$

provided only T_n is consistent. He also shows, under stronger regularity conditions, that the equality holds in (3.52) when T_n is the m.l. estimate of θ. The appearance of Fisher's measure of information in this analysis provides further evidence that this measure is of central importance to estimation.

3.6. *Concluding remarks on efficiency.*

We observe that E_n, for each n, has the maximum local power of discrimination between two neighbouring values (Rao and Poti, 1946) and demand the existence of a statistic T_n independent of θ and having asymptotic correlation unity with Z_n. This ensures that T_n has the same local properties as Z_n. Further it is shown by Wald (1942) that asymptotically shortest confidence intervals can be obtained by inverting regions of the type $Z_n(\theta) = A_n(\theta)$, $Z_n(\theta) < B_n(\theta)$ (one sided regions) and $|Z_n(\theta)| > C_n(\theta)$ (two sided regions). It is clear that any statistic having asymptotic correlation unity with Z_n has the same property in large samples.

It is immaterial what the asymptotic variance of the statistic is provided its asymptotic correlation with Z_n is unity. It may be super or sub efficient in the sense of having smaller or higher asymptotic variance than $[n\ i(\theta)]^{-1}$. If we are placing emphasis on the asymptotic correlation with Z_n being unity we can achieve this by restricting the class of statistics to well-behaved functions of observations. This is for convenience in drawing inferences on θ given the statistic. The m.l. estimate, under some conditions, satisfies our requirements.

Le Cam (1953) suggests asymptotic variance as a measure of concentration of the statistic round the true value in large samples. It may be argued that our interest does not lie in such a measure of concentration. But it is observed that even with respect to such a measure, so far as the existing illustrations suggest, a function of the m.l. estimate serves the purpose.

4. Consistency

A number of quite different examples of inconsistency of m.l. estimates are now available (Neyman and Scott, 1948; Basu, 1955; Kraft and Le Cam, 1956; Kiefer and Wolfowitz, 1956; Bahadur, 1958). The examples have been useful in leading to a proper understanding of the concept of consistency.

Let us consider the concept of consistency as originally introduced by Fisher (1922). We have already referred to it as Fisher consistency (FC) to distinguish it from probability consistency (PC) which figures prominently in statistical literature (ref. Kallianpur and Rao, 1955). A statistic is said to be FC for a parameter θ if

(1) it is an explicit function of the sample distribution function S_n (or the observed proportions $[p_1, ..., p_k]$ in the case of a multinomial), and

(2) the value of the function reduces to θ identically when S_n is replaced by the true distribution function $F(\theta)$, (or the true proportions, $[\pi_1(\theta), ..., \pi_k(\theta)]$ in the case of multinomial).

There is some virtue in such a definition since it is reasonable to demand that the procedure we adopt should give us the true value of the parameter when applied to the entire distribution. Further the definition places some restriction on functions of observations to be considered, whereas in the case of PC there is no restriction on the statistic which can be arbitrary for any finite sample size, however large it may be. So pinning our faith in PC may be somewhat dangerous in many practical situations, where we have to deal with samples of a finite size.

APPARENT ANOMALIES AND IRREGULARITIES IN M. L. ESTIMATION

It is easy to show that an m.l. estimate is FC without any restrictions whatsoever. For if we consider the multinomial situation the log likelihood

$$p_1 \log \pi_1(\theta) + \ldots + p_k \log \pi_k(\theta)$$

when $p_i = \pi_i(\phi)$, is a maximum for $\pi_i(\theta) = \pi_i(\phi)$ which implies, when there is one-to-one correspondence between $\pi_i(\theta)$ and θ, that $\theta = \phi$. In the continuous case the log likelihood

$$\int \log p(x, \theta) \, dS_n$$

has a maximum for $p(x, \phi) = p(x, \theta)$ or $\theta = \phi$ when S_n the sample distribution function is replaced by $F(\phi)$.

What can we say about the m.l. estimate when S_n or $(p_1, \ldots, p_k)$ is close to the true distribution function $F(\phi)$ or $[\pi_1(\phi), \ldots, \pi_n(\phi)]$? We may demand that the distribution in the admissible set maximising the likelihood, $F(\hat\theta)$ or $[\pi_1(\hat\theta), \ldots, \pi_k(\hat\theta)]$, if it exists, should be close to the true distribution. This is true under no condition whatsoever on the admissible class of distributions in the case of a finite multinomial (Hotelling, 1930; Rao, 1957), under the sole condition $\Sigma \pi_i \log \pi_i$ is covergent in the case of the infinite multinomial (Kiefer and Wolfowitz, 1956; Rao, 1958), and under slightly more restrictive conditions in the case of continuous distributions (Wald; 1949; Kraft, 1955). Examples of inconsistency of the estimated distribution functions due to Bahadur (1958), in the cases of an infinite multinomial distribution and a continuous distribution function, show that they are of a very special character, and it appears that it should be possible to prove convergence of the m.l. estimate of the distribution function under fairly weak conditions.

The situation thus appears to be extremely satisfactory so far as the estimated distribution function is concerned. The corresponding convergence in the estimated parameter then takes place when a continuity condition is satisfied, i.e., $F(\theta) \to F(\phi)$ (or $\pi_i(\theta) \to \pi(\phi)$) implies that $\theta \to \phi$. It may be noted that a parameter is, after all, a code number used to identify a distribution and as such it can be arbitrary and need not satisfy any condition. In the examples of inconsistency of m.l. estimates given by Basu (1955)[11] and Kraft and Le Cam (1956) the continuity condition is not satisfied and the examples depend, in a sense, on an unnatural choice of the parameter.

The anomaly regarding inconsistency of the m.l. estimate of a parameter can be resolved to some extent if we consider consistency in a broader sense as mentioned in Section 2 of this paper. It was observed that if for any two given values of the parameter the distributions of the observations tend to be orthogonal as the sample size $\to \infty$, it is reasonable to demand that the distributions of the estimate also behave in the same way. When this is so we may say the estimate is consistent for the parameter in the wide sense. Such wider consistency is ensured when the estimate tends to two different constants for two different

---

[11] Basu (1955) gave the example of a binomial distribution where the probability of success $p(\theta)$ is defined as follows :

$$p(\theta) = \theta \qquad \text{if } \theta \text{ is rational}$$
$$= 1 - \theta \quad \text{if } \theta \text{ is algebraic irrational}$$

The m.l. estimate of θ, which is the observed proportion of success, tends to θ when θ is rational and to $(1 - \theta)$ when θ is algebraic irrational, and is thus not consistent. Basu also shows that there exists another estimate which is consistent for θ. The example of Kraft and Le Cam is more complicated.

values of the parameter. We need not insist that the constant to which the estimate tends should be equal to the true value of the parameter. In Basu's example, (footnote 11) the m.l. estimate tends to θ when θ is rational and to $(1-\theta)$ when θ is algebraic irrational; thus the m.l. estimate is consistent in the wide sense. The same is true of the example[12] considered by Neyman and Scott (1948) where the m.l. estimate of σ^2, the structural parameter, tends to $(n-1)\,\sigma^2/n$ and not to exactly σ^2; clearly, the m.l. estimate is consistent in our sense. The m.l. estimate is, however, not consistent even in the wide sense in Bahadur's examples.

We may also consider a slightly different kind of example due to H.E. Daniels (quoted in a paper by Kendall and Babington Smith, 1950). Observations $(x_1, y_1), ..., (x_n, y_n)$ are such that

$$x_i = \alpha_i + \epsilon_i, \qquad y_i = \alpha_i + \mu + \eta_i$$

where ϵ_i is $N(0, \sigma^2)$, η_i is $N(0, \zeta^2)$, and ϵ_i and η_i are independently distributed. Simultaneous estimation of α_i, μ, σ^2 and ζ^2 by the m.l. method leads to the same value for the estimates of σ^2 and ζ^2, so that the estimates of these two parameters are clearly inconsistent in any sense. This result is perhaps not surprising, for the data *themselves* do not seem to provide satisfactory discrimination between σ and ζ (or between one pair of values of σ, ζ and another pair) however large may be number of pairs of observations, when nothing is known about the behaviour of the incidental parameters α, as $i \to \infty$.

5. Conclusion

Since the main aim of this paper is to consider apparent anomalies in the m.l. method, no reference has been made to the superiority of the m.l. method over others. It may be claimed that certain other methods also provide estimates which have the same properties as the m.l. estimates in large samples, although they may be subject to similar criticism in other respects. This may be true, but we cannot use asymptotic properties as sole criteria for the selection of a technique which has to be applied in finite samples in practice. So we have to look for other properties, which hold good for all sample sizes. We may list here some properties of this type which support the claims of m.l. estimates.

The m.l. method has wide applicability. The m.l. estimate is a function of a minimal sufficient statistic, and in special cases is itself a minimal sufficient statistic, a property which may be considered desirable (ref. Rao, 1945, 1946, 1948) and which is not shared, in general, by other general methods of estimation. Finally, consideration of the likelihood function enables us to recognise the minimal sufficient statistic, and if necessary, to supplement the m.l. estimate with other statistics to recover part of the information lost in using the m.l. estimate alone. A more accurate measure of loss of information, based on the variance of Z_n given an estimate T_n (Fisher, 1925), the asymptotic value of which is more appropriate for comparing statistics when the sample size is not very large, shows that the loss associated with the m.l. estimate is smaller when compared to many other procedures. A detailed study of this aspect is undertaken in Rao (1960).

---

[12]Neyman and Scott (1948) consider an increasing sequence of s series of measurements x_{ij} ($i = 1, 2, ...,$ $s \to \infty$, $j = 1, ..., n$), all independently distributed. The probability law of x_{ij} is normal with mean α_i and variance σ^2. The parameter σ^2 is called structural and is the same for all observations, while the α_i, which vary from series to series, are called incidental parameters. The maximum likelihood estimate of σ^2 is $\Sigma \Sigma (x_{ij} - \bar{x}_i)^2/sn$ which is consistent for $(n-1)\,\sigma^2/n$ and not for σ^2.

APPARENT ANOMALIES AND IRREGULARITIES IN M. L. ESTIMATION

REFERENCES

BAHADUR, R. R. (1958) : Examples of inconsistency of maximum likelihood estimates. *Sankhyā*, **20**, 207.

——— (1960) : On the asymptotic efficiency of tests and estimates. *Sankhyā*, **22**, 229.

BARANKIN, E. W. and Gurland, J. (1951) : On asymptotically normal and efficient estimates. *Univ. California Publ. in Stat.*, **1**, 89.

BHATTACHARYA, A. (1946) : On a measure of divergence of two multinomial populations. *Sankhyā*, **7**, 401.

BASU, D. (1952) : An example of non-existence of minimum variance estimator. *Sankhyā*, **12**, 43.

——— (1954) : Choosing between two simple hypotheses and the criterion of consistency. A chapter in the D. Phil thesis submitted to the Calcutta University.

——— (1955): An inconsistency of the method of maximum likelihood. *Ann. Math. Stat.*, **26**, 144.

——— (1956) : The concept of asymptotic efficiency. *Sankhyā*, **17**, 193.

BERKSON, J. (1955) : Estimation by least squares and by maximum likelihood. *Proc. Third Berkeley Symposium on Mathematical Statistics and Probability*, **1**.

CRAMER, H. (1946) : *Mathematical Methods of Statistics*, Princeton University Press.

DOOB, J. (1934) : Probability and Statistics. *Trans. Am. Math. Soc.*, **36**, 766.

——— (1936) : Statistical estimation. *Trans. Am. Math. Soc.*, **39**, 410.

FISHER, R. A. (1922) : On the mathematical foundations of theoretical statistics. *Phil. Trans. Roy. Soc. London*, Series A, **222**, 309.

——— (1925) : Theory of statistical estimation. *Proc. Cambridge Phil. Soc.*, **22**, 700.

HELLINGER, H. (1909) : Neue Begründung der Theorie quadratischer Formen von unendlich-vielen veränderlichen. *J. für. reine und angew, Mathematik*, **136**, 210.

HOTELLING, H. (1930) : The consistency and ultimate distribution of optimum statistics. *Trans. Am. Math. Soc.*, **32**, 847.

HUZURBAZAR, V. S. (1948) : The likelihood equation, consistency and the maximum of the likelihood function. *Ann. Eugenics*, **14**, 185.

KALLIANPUR, G. and RAO, C. R. (1955) : On Fisher's lower bound to asymptotic variance. *Sankhyā*, **15**, 331.

KENDALL, M. G. and SMITH, B. BABINGTON (1950) : Factor analysis. *J. Roy. Stat. Soc.*, Series B (Methodological) **12**, 60.

KIEFER, J. and WOLFOWITZ, J. (1956) : Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Ann. Math. Stat.*, **27**, 887.

KRAFT, C (1955) : Same conditions for consistency and uniform consistency of statistical procedures. *Univ. California Publications in Statistics*, **2**, 125.

KRAFT, C. and LE CAM, L. (1956) : A remark on the roots of the maximum likelihood equation. *Ann. Math. Stat.*, **27**, 1174.

KULLBACK, S. and LIEBLER, R. A. (1951) : On information and sufficiency. *Ann Math. Stat.*, **22**, 79.

LE CAM, L. (1953) : On some asymptotic properties of maximum likelihood estimates and related Bayes' estimates. *Univ. of California Publ. in Stat.*, **1**, 277.

LEHMANN, E. and SCHEFFÉ, H. (1950) : Completeness, similar regions and unbiassed estimation. *Sankhyā*, **10**, 305.

NEYMAN, J. (1949) : Contributions to the theory of the χ^2 test. *First Berkeley Symposium on Mathematical Statistics and Probability*, 239.

NEYMAN, J. and SCOTT, E. L. (1948) : Consistent estimates based on partially consistent observations. *Econometrica*, **16**, 1.

RAO, C. R. (1945) : Information and accuracy attainable in the estimation of statistical parameters. *Bull. Cal. Math. Soc.*, **37**, 81.

——— (1946) : Minimum variance and the estimation of several parameters. *Proc. Cambridge Phil. Soc.*, **43**, 280.

——— (1948) : Sufficient statistics and minimum variance estimates. *Proc. Cambridge Phil. Soc.*, **45**, 215.

——— (1957) : Maximum likelihood estimation for the multinomial distribution. *Sankhyā*, **18**, 139.

——— (1958) : Maximum likelihood estimation for the multinomial distribution with infinite number of cells. *Sankhyā*, **20**, 211.

——— (1960) : Asymptotic efficiency and limiting information (to appear in the *Proceedings of the Fourth Berkeley Symposium on Probability and Statistics*).

Rao, C. R. and Poti, S. J. (1946) : On locally most powerful tests when alternatives are one-sided. *Sankhyā*, **7**, 439.

Silverstone, H. (1957) : Estimating the logistic curve. *J.A.S.A.*, **52**, 567.

Wald, A. (1942) : Asymptotically shortest confidence intervals. *Ann. Math. Stat.*, **13**, 127.

———— (1943) : Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Trans. Amer. Math. Soc.*, **54**, 426.

———— (1949) : A note on the consistency of the maximum likelihood estimation. *Ann. Math. Stat.*, **20**, 595.

Résumé

La méthode de maximum de vraisemblance (m.l.) pour l'estimation des paramètres inconnus a été censuré pour les raisons suivantes : (i) elle ne donne pas un estimateur convergent et (ii) il y a des estimateurs plus efficients que ceux de maximum de vraisemblance et (iii) l'usage du m.l. entrainet de la computation difficile. Cette dernière critique ne serait pas importante si les machines electroniques capable d'accepter les instructions compliquées concernant des opérations numériques, se font disponibles chez travailleurs.

L'on indique ici, d'abord, que le but de l'estimation est la condensation des données sans perte de l'information essentielle et puis l'on fournit une certaine justification pour la mesure de l'information de Fisher et le critère de maximisation de l'information dans une statistique. Les conceptions de l'efficience et de la consistence ont été reformulées afin que l'on puisse fournir un critère pour une telle choix de l'estimateur que la perte de l'information donnée soit negligible dans les grands échantillons. Un estimateur efficient a été défini comme une estimateur qui a une corrélation asymptotique de mesure d'unité avec la derivé du logarithme de la vraisemblance. Un estimateur de maximum de vraisemblance sous quelques conditions, est efficient dans ce sens.

L'équivalence de cette définition avec celle de Fisher qui constate que la consistence est l'atteinte de moindre variance asymptotique, est établie dans quelques conditions de régularité sur la statistique. Mais la définition nouvelle résout la difficulté qui a apparué grâce à l'existence des estimateurs super-efficients ayant, peutêtre, une variance asymptotique plus petite que la variance auprès des estimateurs de maximum de vraisemblance.

L'on montre ici, que les estimateurs super-efficients sont équivalents aux estimateurs de maximum de vraisemblance (quand ils sont efficients dans ce sens nouveau) ou sont inférieurs auprès des estimateurs de maximum de vraisemblance pour servir le but de l'inférence statistique, (quand ils ne sont pas efficients dans le sens nouveau).

On dit qu'un estimateur soit consistent dans le sens plus ample si ses distributions asymptotiques pour deux valeurs différentes du paramètre, soient orthogonales. Plusieurs exemples de l'inconsistence des estimateurs de maximum de vraisemblance dans le sens ordinaire paraissent remplir la condition de la consistence plus étendue.

L'on indique ici la consistence de la fonction estimée de distribution est plus fondamentale que celle de l'estimateur du paramètre particulier. Cette dernière consistence suive naturellement si le paramètre est défini comme un fonctionnel continu de la fonction de distribution. Mais un paramètre est rien qu'un nombre de code défini pour l'identification d'une distribution et par conséquence, le paramètre ne doive remplir la condition posée. Les irrégularités à l'ègard de l'inconsistence de l'estimateur de maximum de vraisemblance laissent s'expliquer par le défaut de cette condition.

Cette étude conclut avec une note sur les propriétés essentielles des estimateurs de maximum de vraisemblance en cadre des échantillons petits.

APPARENT ANOMALIES AND IRREGULARITIES IN MAXIMUM LIKELIHOOD ESTIMATION

PRESIDENT : E. J. G. PITMAN

1. Apparent anomalies and irregularities in maximum likelihood estimation

L'auteur, M. Rao, présente sa communication[1]

MR. NEYMAN : 1. Mr. Rao's very interesting paper brings out certain philosophical questions regarding criticisms levelled at maximum likelihood estimation and, in addition, presents an extensive history of the problem going back to 1922 when the term Maximum Likelihood Estimate (MLE) was first used. The purpose of the present note is to contribute to both subjects : to express my views on the philosophy of theoretical statistical research and to push the historical sketch back to 1908 when the ideas or certain properties of the MLE seem to have been first expressed.

2. The philosophical aspect of the problem is connected with the two different points of view on statistics, one having to do with intensities of belief and the other behavioristic. To me personally[2] the intensity-of-belief theory of statistics appears dogmatic and, as reflected in the writings of the various authors, is reducible to proposals, occasionally quite insistent proposals, to adopt specified formulas as measures of intensities of belief which an individual should experience in specified circumstances. One such theory, or creed, advises special formulas as *a priori* probability distributions for unknown parameters to be used in cases where the circumstances of the problem do not imply specific *a priori* distributions or even do not imply that the parameter considered is a random variable. Further advice is to use the recommended *a priori* distribution for substitution in the familiar Bayes' formula.

Another modification of essentially the same dogmatic school of thought is based on the premise that the concept of probability is a measure of intensity of belief which is applicable in some cases but not in all. For these cases where the probability is not applicable to measure the uncertainty, the proponents of the relevant school of thought devise new measures of confidence or diffidence and one of them is the mathematical likelihood. In thinking of these and similar attempts at foundations of mathematical statistics, I recall the expressive title of two articles of our recently deceased colleague and friend, D. van Dantzig : "Statistical Priesthood" I and II.[3]

The alternative point of view on foundations of statistics, the behavioristic or operational point of view, stems from some ideas of Laplace and, expressed somewhat more clearly, of Gauss. Leaving aside the question of confidence and diffidence, the behavioristic point of view concentrates on those cases where the mathematical probability is an idealization of relative frequencies as experienced in the realm of natural phenomena. Here, as is most frequently the case, we are confronted with the necessity of a choice among a number of possible actions and the desirability of each action depends upon the value of a parameter intervening in the distribution of the observable random variables. If the value of this parameter is known or assumed known, there is no problem. Statistical problems arise when the relevant parameter is not known and the choice of action has to be based on the values of the observable random variables, that is, on the value of an estimator of the parameter. The problem of estimation is, then, to devise the estimator. This problem splits into a number of detailed problems. One is to establish the properties of all the different estimators, that are available to choose from in a given problem. Another detailed problem is to devise the method, just as easy a method as possible, of calculating the estimator having the properties that fit the situation best.

The properties of an estimator which may be considered desirable vary from one particular problem to the next. Also undoubtedly, they depend on subjective elements : it is quite conceivable that two different persons contemplating the same problem will have different preferences for the properties that an estimator should have. In some problems and to some individuals, unbiasedness of the estimator and the smallness of its variance appear of paramount importance. Here, Gauss' method of least squares is frequenlty the answer. In other cases, unbiasedness and small variance are secondary or irrelevant, and some other property appears important. Thus, for example, in the problem of estimating the degree of contamination of drinking water, it may appear most important not to underestimate the contamination and, from the point of view of public health, the most desirable estimator is certainly not unbiased.

---

[1] *Bull. Inst. Int. Stat.*, XXXVIII, 4, p. 439.

[2] J. Neyman, "Inductive behaviour as a basic concept of philosophy of science." *Rev. Int. Stat. Inst.*, Vol. 25 (1957), pp. 7–22.

[3] D. van Dantzig, "Statistical Priesthood" I and II, *Statistica* Neerlandica, Vol. II (1957) and Vol. 12 (1958).

With reference to Rao's discussion of criticisms of maximum likelihood estimation, I wish to make it clear that my own criticisms are directed not towards the use of MLE but to the insistence that these estimators, or indeed any other estimators, be used as a matter of principle. In my opinion, any user of statistical methods should have complete freedom of choice and the role of theory in the matter is to elucidate the properties of the methods that are available. Thus, for example, the famous inequality giving the greatest lower bound of the variance of an unbiased estimator, first found by Fréchet and then, independently, by Harald Cramer and Rao, is a very important result. It is purely behavioristic or operational and tells us that, under certain conditions, if we insist on using unbiased estimators then, no matter what we do, the variance of the estimator cannot be less than a calculable limit. As indicated by a slightly more general version of the same inequality, there may be a possibility of finding an estimator which has mean square error less than the bound of Rao; then this estimator must be biased. With this in mind, the consumer of statistical theory may perhaps decide to drop the requirement of unbiasedness.

For quite some time the possibility of biased estimators with mean square errors less than the bound for the variance of an unbiased estimator remained just a theoretical possibility with no live example to show that they really exist. Then Mr. Joseph Berkson appeared on the scene and produced a real case of an estimator, obtained by minimizing the classical Karl Pearson χ^2, which not only has its mean square error less than that of MLE but also less than the indicated bound! I submit that this particular result of Berkson is of considerable interest and importance. Quite apart from the possibility of estimating the parameter with precision, in the sense of mean square error, better than other known estimators, this result raises a host of novel theoretical problems: what are the situations in which biased estimators exist with their mean square errors less than the Fréchet-Cramér-Rao lower bound for variances of unbiased estimators? Can one invent a method, a machinery such as the maximalization of the likelihood or the minimalization of the χ^2, which, at least in some cases, would lead to such estimators if they exist ?

I note that Rao does not particularly like Berkson's result, apparently for the reason that in the particular problem considered, Rao's own interest centers on a parameter different from the one estimated by Berkson. Evidently, we consider the question from different points of view.

3. Turning to the other part of my contribution, concerned with a detail in the history of MLE I find it interesting that, at least on two occasions, the idea of MLE sprang up from the dogmatic intensity-of-belief approach to statistics. However, in both cases the original dogmatic approach was followed by studies of a distinctly behavioristic or operational character. The two approaches can be roughly summarized as follows :

(i) *First statement :* MLE should be used because this use is implied by such and such principle.

(ii) *Second statement :* The consistent use of MLE will guarantee such and such long range advantages.

As far as I am aware, the priority in the approach to MLE just described, involving both statements (i) and (ii), belongs to F.Y. Edgeworth.[4] The dogmatic intensity-of-belief ideas of Edgeworth, which are also noticeable in Laplace, were connected with the arbitrary *a priori* distributions and the use of Bayes' formula. The fact that this brought Edgeworth to the use of what we now call MLE is occasionally noted in the literature. For example, an appropriate reference is found in M.G. Kendall's book.[5] However, it is much less generally known and seems to have escaped the attention of Rao, that, after making statements roughly equivalent to (i), Edgeworth proceeded to formulate a conjecture in the spirit of the statement (ii) above. The passage I have particularly in mind, published in 1908, is printed in the Appendix of a very large and involved paper. It so happens that this conjecture of Edgeworth is now known to be broadly true but with some exceptions. Also, as reflected in the excellent historical summary given in the present paper by Rao, although 52 years have elapsed since the publication of Edgeworth's conjecture, the limits of its validity are still the subject of numerous studies all over the world. In these circumstances and because of the general lack of awareness of the identity of the author of the conjecture, it appears appropriate to reproduce here a brief quotation from the Appendix of Edgeworth's paper.

---

[4] F. Y. Edgeworth, "On the probable error of frequency constants" *J.R.S.S.*, Vol. 71 (1908), pp. 651- 678.

[5] M. G. Kendall, *"The Advanced Theory of Statistics"* Vol. II, Griffin, London, 1946.

APPARENT ANOMALIES AND IRREGULARITIES IN M. L. ESTIMATION

Appendix

This Appendix is designed as a receptacle for some mathematical developments which might have interrupted the course of the preceding arguments.

I *Love's proof of some preceding propositions.*

A foremost place is due to Love's confirmation of certain propositions above stated by means of an independent proof.

The first proposition thus verified is a particular case of a general theorem which may thus be provisionally restated. Let $y = e\psi(x)$, be a frequency-function apt to represent the distribution of statistical observations. Let $x_1, x_2, ..., x_n$ be a set of n observations forming a random selection from the indefinitely large group of the observations ranging under the frequency curve. Let $\phi(x_1, x_2, ..., x_n)$ be that function of the given observations, which affords the *most probable* value (as determined by inverse probability) of the sought point to which the observations relate; a symmetrical function when, as will be here supposed, the observations are all of equal weight or worth. Then, if we take (at random) a series of sets, such as

$$\begin{matrix} {}_1x_1 & {}_1x_2 & ..., & {}_1x_n, \\ {}_2x_1 & {}_2x_2 & ..., & {}_2x_n, \\ . & . & . \\ . & . & . \\ . & . & . \\ {}_mx_1 & {}_mx_2 & ..., & {}_mx_n; \end{matrix}$$

and form for each set the corresponding value of ϕ, the series of mean values thus formed say, ${}_1\phi, {}_2\phi, ..., {}_m\phi$ will be such that (m and n being large numbers) the mean square of their deviation from the true point, say x, viz.,

$$\frac{({}_1\phi-x)^2+({}_2\phi-x)^2+\cdots+({}_m\phi-x)^2}{m}$$

will be less than the mean square of deviation presented by any other set of mean values ${}_1\chi, {}_2\chi, ..., {}_m\chi$, each formed from a set of n observations, where χ (like ϕ) is a symmetrical function of observations, having the properties of an average.

In line with the Victorian style, the above passage is interspersed with footnotes. I take the liberty of omitting these.

In contemplating this passage, one is struck by the change in style, terminology and precision of expression which have occurred during the half-century that elapsed since the publication of Edgeworth's paper. However, the translation of the passage into modern terms presents little difficulty.

The function $y = \exp\{\psi(x)\}$ is the probability density of an observable random variable, say X. Further context suggests that, in addition to x, the particular value of X, the function ψ depends upon a parameter, say θ, and that the actual value of this parameter, say θ_0, is unknown. The value θ_0 is described by Edgeworth as "the sought point to which the observations relate" and, later on, as "the true point, say x, ...". In order to avoid the use of the same letter x in several different meanings, I introduce the symbol θ_0.

The symbols ${}_ix_j$ for $i = 1, 2, ..., m$ and $j = 1, 2, ..., n$ represent independent observations on the variable X arranged in m samples of n observations each. The function ϕ described as "that function of the given observations, which affords the *most probable* value (as determined by inverse probability)" is simply the maximum likelihood estimator of θ_0.

Edgeworth's assertion is that, if X is an alternative estimator of θ_0, based on the same observations as ϕ and subject to some not distinctly stated limitations: "where χ (like ϕ) is a symmetrical function having the properties of an average," then the asymptotic mean square error of ϕ will be less than (presumably, not greater than) that of χ.

Edgeworth was not able to prove his conjecture to his own satisfaction and tried to enlist the help of Love. Unfortunately, Love's success was limited. However, Edgeworth's assertion compares favourably with those found in a number of recent books on statistics which flatly assert that, as proved by somebody or other, the asymptotic variance of MLE is a minimum, without any limitations. In favour of Edgeworth is the realization that some sort of restriction on the alternative estimator χ is necessary.

The ideas of Edgeworth did not seem to have much influence on the thinking of the contemporary statisticians, and the above clear cut statement of the presumed optimal property of MLE went unnoticed.

12

The idea reappeared in the literature fourteen years later in the famous paper of R.A. Fisher to whom we owe a great number of other concepts and terms, consistency, efficiency, sufficiency, etc. Here, again, the origin of MLE was in the degree-of-belief approach to statistics but based on principles different from those of Edgeworth. Also in this case, the original approach, which appears to me dogmatic, was followed by increasingly accurate behavioristic studies. This part of history appears adequately covered by Rao, and I need not enter in any details.

4. Before concluding I would like to request Professor Rao to explain his philosophical standpoint a little more clearly than he does in his paper. Some passages in his paper suggest the possibility that, in Rao's opinion, MLE should always be used irrespective of the properties it may have. Does Rao really mean this? The passages I have in mind include Rao's *Introduction* and his *Conclusions*.

In his Introduction Rao lists four different reasons for which maximum likelihood estimation has been mainly criticised. In the Conclusions there are listed the various advantages of MLE. It looks as if the choice of a method of estimation is treated more or less like the choice of an automobile which a family will have to use for a number of years. All automobiles on the market are open to some criticisms and some of them have certain advantages. The standpoint of Rao seems to be that the advantages of the automobile MLE outweigh the disadvantages. This impression is fortified by Rao's dealing with what he considers as criticisms of MLE. One reason listed is that in certain cases MLE have been shown to be inconsistent in the usual sense of the term. This fact is not denied by Rao. Instead, he introduces a distinction between PC and FC, consistency in the sense of convergence in probability and consistency in the sense of Fisher. Also there are some other interesting connections in which the term consistency is used. It is then shown that in some cases where the MLE are inconsistent in one sense, they are consistent in another sense.

Another ground for criticisms discussed by Rao is that, in some specified cases, consistent estimators of a parameter θ are readily available with mean square errors that are less than those of MLE. In one such case, Rao's stand seems to be that it is pointless to try to estimate θ. In my opinion, the difference between selecting an automobile and selecting a method of estimation is that the car is, so to speak, indivisible. It is impossible for a purchaser to take some characteristics of a Volkswagen, very desirable for short trips in town, and combine them with certain other characteristics of a Rolls Royce, most desirable for extensive travel. If the family is limited to a single car, it must face the necessity of weighing the relative advantages of each make against the disadvantages. No such necessity exists in the choice of a method of estimation. Provided one knows the properties of the several methods available for the given problems, and provided one is clear as to what one wants to achieve, one is at liberty to use MLE in certain cases and some alternative estimators in others. However, in order to avoid disappointments, it is quite essential to know what the properties of the different estimators are.

From this point of view, the authors whom Rao considers as critics of MLE, are not really critics. They just provide us with valuable information.

In order to make Rao's philosophical stand quite clear, I suggest that he gives an unambiguous answer to a trivial question which, however, is both specific and illustrative. Suppose that an association of manufacturers of certain measuring instruments is anxious to have a formula for estimating the error variance σ^2 of each instrument. Suppose that with each instrument a moderate number n of independent measurements are made of a large number N of different objects. With the usual assumptions and with the usual notation, the two contemplated estimators are

$$S_1^2 = \frac{1}{nN} \sum_{i=1}^{N} \sum_{j=1}^{n} (x_{ij} - x_{i\cdot})^2$$

and

$$S_2^2 = n S_1^2 / (n-1).$$

The first estimator S_1^2 is the ML estimator. The second is not. However, the first estimator has an operational property which may seem undesirable: it is inconsistent. In fact, as $N \to \infty$, the first estimator tends in probability not to σ but to a smaller number $(n-1)/n\sigma^2$. On the other hand, the second estimator is consistent and, in fact, unbiased.

The question is: which of the two estimators would Rao recommend? I hope that Rao's advice will be behavioristic, in favour of S_2^2. If it is not and if he insists on MLE, there may be trouble. In fact, there may be a lawsuit for damages. For, if one of the manufacturers, say A, has his $n = 10$ and another manufacturer B has $n = 2$, the manufacturer A will have a legitimate reason to complain.

APPARENT ANOMALIES AND IRREGULARITIES IN M. L. ESTIMATION

Here is another question. In his paper Rao writes that someone has suggested that the super efficient estimators constructed by Hodges and Le Cam be used in practice. Would Rao kindly indicate who made this suggestion. The point is that, with Rao's statement as it now stands, the reader is likely to think that the suggestion came from either Hodges or Le Cam or from both. At this I would be most surprised.

The Hodges-Le Cam estimators are the usual count examples showing that certain theorems, thought to have been proved rigorously, are in fact, false. In the present case the theorem in question is : out of all consistent and asymptotically normal estimators, the MLE has a minimum asymptotic variance for all values of the estimated parameter. Hodges' example showed that this theorem is false and that, for the assertion to be true, it is necessary to consider not all the estimators of the kind described but those of some limited class.

Résumé

Le papier intéressant de M. Rao me suggère quelques réflections d'ordre philosophique et quelques autres d'ordre historique. Premièrement, il me parait frappant que les estimateurs de "Maximum Likelihood" (M.L.E.) semblent être recommandés par certains auteurs, dont M. Rao, pour des raisons de deux espèces différentes : parce que dans certains cas ces estimateurs possèdent des propriétés désirables et parce que leur usage systématique est une affaire de principe, indépendamment des conséqnences. La première raison est toute naturelle, mais la deuxième me paraît étrange. Voici un exemple. Considérons quelques instruments à mesurer. Pour charactériser leur précision, on emploie chaque instrument pour faire m mesures indépendantes sur chacun des n objets différents. Soit $X_{ij} = \mathcal{N}(\xi_i, \sigma^2)$ une de ces mesures. Une agence publique, dont le but est de charactériser la précision moyenne des instruments produits par différentes fabriques, a besoin d'un estimateur de la variance σ^2 de l'erreur de mesure. La formule

$$S^2 = \sum_{i=1}^{n} \sum_{j=1}^{m} (X_{ij} - X_i.)^2 / mn \tag{1}$$

représsute l'estimateur M. L. Supposons que deux usines, A et B, produisent des instruments identiques, avec $\sigma = 1$. Supposons que dans la fabrique A on a $m = 2$. Alors, comme on le sait bien, lorsque n augmente, $\lim p S^2 = 0.5$, D'autre part, si dans la fabrique B on a $m = 10$, alors $\lim p S^2 = 0.9$. Donc, dans ce cas, l'application de M.L.E. conduirait à une conclusion fausse que les instruments venant de A sont beaucoup plus précis que ceux venant de B. D'autre part, il est aisé de définir un estimateur de σ^2 n'ayant pas cet inconvenient. Ce qui m'intéresse c'est si M. Rao recommandrait l'usage de (1), meme dans les conditions indiquées, pour l'unique raison que cet usage est prescrit par le principe de M. L. de M. Fisher.—Un détail historique : à ma connaissance, la première tentative de formuler un théorème impliquant les propriétés désirables des M. L. E. se trouve dans un travail de F. Y. Edgeworth publié en 1908. J'en cite un passage dans mon texte anglais.

Mr. Kitagawa : I believe that, Mr. Rao has been most successful in attaining his main purpose in this paper, namely, in resolving apparent anomalies and irregularities in maximum likelihood by reformulating the notions of efficiency and consistency in a very natural and elegant way and also in closer connection with the original ideas of Sir Ronald Fisher. It is the merit of the present paper that further discussions can and must be done from any more essential standpoints includ'ng philosophical ones than those connected merely with mathematical techniques.

Mr. Berkson : Professor Rao has presented a novel concept of estimation, and it is interesting to visualize its operation in the bio-assay case. Recall how the bio-assay problem arises in its medical application. The physician has ordered the administration of, say, 200 units of insulin in a unit volume and the pharmacist prepare a solution with that concentration. If his stock solution contains 400 units per unit volume, he will dilute it to half its strength. If it contains m units he will dilute it in a proportion of $1/m$. He makes a bio-assay to find out what is the value of m. In terms of the decision concept of estimation, the decision involved here is the number of cubic centimeters of diluent to add to the stock solution, in order to bring it to the strength required by the physician. The measure of the efficacy with which the bio-assay is accomplished is some average of the error made in estimating m, and the classic measure used, though not the only conceivable one, is the mean square error. Perhaps we will give it loftier statistical prestige if we call it a "loss function."

Mr. Rao suggests that in statistics, estimation is only an incidental procedure—that the serious statistical objective is the condensation of the data to economical form so that, for instance, they may be added to similar data obtained in another bio-assay. We can imagine doing this. We perform a bio-assay and record the results in a statistically efficient way, perhaps as a minimal sufficient statistic, perhaps as the whole likelihood function. When we perform another bio-assay we will, in an efficient manner, add the efficiently summarized data of that bio-assay to the efficiently summarized data of the first bio-assay. When we make another bio-assay, we will add the efficiently summarized data to those already accumulated, and so forth. Now, if instead of making a point estimate in the first instance and diluting the solution to the required 200 units in accord with that point estimate, the statistician faithfully pursues the avowed purpose of estimation to condense the data and to prepare a whole series of fiducial limits in the sense of Fisher, or confidence limits in the sense of Neyman, what will happen ? Well, what will happen in the meantime is that the patient will die in diabetic coma ! This, of course, is irrelevant to the logical development of the fundamentals of statistics, but it is a point. Another point is that the law, in its benightedness, does not allow the killing of a patient with an overdose of inference theory and an underdose of insulin. If I am the statistical bio-assayist following Rao's theories –and this is conceivable since I am a great admirer of Rao—I will be committed to the hoosegow on the charge of malpractice. I hope that while I am in durance vile, my friend Rao will visit me. It will be a consolation to contemplate with him the ultimate nature of statistics and to realize that while I am suffering on bread and water it is in the noble cause of statistics considered as right thinking and correct rational inference, regardless of practical consequences.

Now a word about the summarization of data. Suppose we accept Fisher's measure of efficiency as the proportion of available information (in a certain sense), extracted by a statistic T [8].[6] It is obvious that this cannot be a measure of the efficacy of T as an estimator. Any "random" number or even a meaningless symbol T that is a one-to-one function of the possible samples will be completely efficient (sufficient) in this sense. However, it would hardly do as an estimator. But considering Fisher's efficiency only as a measure of effective condensation of data, what is the relation of it to maximum likelihood estimation ? It should be emphasized that Rao definitely did not say that this estimator necessarily extracts as much information as possible. But I have the impression that such a claim has been made, though this may be a misunderstanding. It seems to be widely believed for instance that where a sufficient statistic other than the sample exists, the maximum likelihood estimate will be sufficient and hence will extract the total amount of information available [8] [9] [10] [11]. But this is not strictly true. What seems to be true is that the maximum likelihood estimate will be a function of the minimal sufficient statistic, but it will not necessarily be a one-to-one function, and therefore it will not necessarily be sufficient. In such cases—and they seem to be of fairly common occurrence, e.g. [5]—the maximum likelihood estimate would not be sufficient even for storing the total "information." For this purpose, one should have to store at least the sufficient statistics themselves. In the instance of the logistic function with binomial variation, there are minimal sufficient statistics for its parameters α, β. For the case of a "bio-assay" experiment with three equally spaced "doses" x, $n = 10$ animals exposed at each dose, both parameters to be estimated, a minimum χ^2 estimate which I call the "minimum logit χ^2 estimate" is consistent FC as well as consistent PC, and is asymptotically efficient. For finite samples it is sufficient, and extracts the total amount of available information. The same is true for an infinite number of other controllable experimental arguments, though not in all such arrangements. The maximum likelihood estimate is also consistent FC and PC and is asymptotically efficient, but for finite samples it has larger mean square error than the minimum logit χ^2 estimate, and it is not sufficient. It loses a calculable amount of information, which is small for an experiment in which the probability P_c of response at the central dose is 50 per cent, but the proportion of information lost increases as the experiment is asymmetrically placed, and approaches unity as P_c approaches 1 or zero. I should like to ask Professor Rao whether, with an experiment such as described, he would still prefer the maximum likelihood estimator to the minimum logit χ^2 estimator.

---

[6]Rao reiterates this definition, but one should note that Fisher did not limit it to asymptotically normal estimators, and specifically applied it "to finite samples and to other cases where the distribution is not normal." The definition is pertinent as a measure of the sufficiency of a statistic, but not as the efficiency of an estimator. This distinction is widely recognized.

APPARENT ANOMALIES AND IRREGULARITIES IN M. L. ESTIMATION

My position on estimation is crudely simple to the point of simple-mindedness. Statistics is used for many purposes. One of these, and in my own opinion it is the centrally important one, is to estimate a specified parameter. When this is the objective, the measure of the relative worth of an estimator is the value of some loss function such as the mean square error. In using the mean square error, I do not mean to make the world believe that its loss is always proportional to the square of its error, as Rao implies, but take it only as a representative loss function though, of course, it is the classic measure of error and the most widely used. It does seem to be a good working rule to suppose that the probability of making an error greater than a critical size is probably larger with increase of the error variance of the measure, though the conditions in which this is certainly true are limited indeed. I do not know any theorems in which the probability of significant errors is smaller, the larger the mean error, but I know some in the opposite sense, for instance the normal law as a precise case, and the Tschebysheff rule for an approximate evaluation. I do not think that it is a matter of indifference whether the mean square error of an estimate is large or small, and I see no reason for preferring an estimate with large mean square error. Knowing nothing relevant to the contrary about an assay, I should want medicine that I prescribed to be assayed by a method with known small mean error. Indeed, I should feel duty bound to insist on it. And if there were a reason for my disregarding the small mean error, it could not be because there was another method that better condensed the data. I quite disagree with Rao when he defines the purpose of estimation as the condensation of data. The object of estimation is to evaluate the parameter with as little error as possible, in some acceptable definition of "error." To define the objective as condensation of data, irrespective of error, seems to me not to point up the essential purpose of estimation, but to divert us from it. Rao's apparent predilection for an assay with large average error seems to me unnatural. His present belittling of small mean square error is puzzling. I notice that elsewhere he characterizes an unbiased estimate as "best" if it has minimum attainable variance [12].[7]

Now this does not mean that the loss function of mean square error is the only conceivable one, or that it is necessarily definitive. If, in a particular application, some other loss function suggests itself, let it be investigated. Rao has questioned my use of the mean square error, which is the loss function of Gauss,[8] when comparing some minimum χ^2 estimates with the maximum likelihood estimate of the parameters of the logistic function and of the integrated normal function. In these investigations it was found

---

[7] Rao has informed me that he was only using accepted terminology here, without implying that such an estimator is best from a practical view. Even so, the use reflects a generally accepted attitude and does not support Rao's suggestion that his view is shared by most statisticians.

[8] Since Edgeworth and Gauss have been mentioned in this discussion, the following quotation from Edgeworth [7] referring to Gauss is interesting:

The reflections of the great mathematician on this branch of mathematical physics deserve to be transcribed here —"That the metaphysic employed in my Theoria Motus Corp. Coel...to justify the method of least squares has been subsequently allowed by me to drop (Dass ich....habe fallen lassen) has occurred chiefly for a reason that I have myself not mentioned publicly. The fact is, I cannot but think it in every way less important to ascertain that value of an unknown magnitude the probability of which is the greatest—which probability is nevertheless infinitely small —rather than that value by employing which we render the Expectation of detriment a minimum (an welchen sich haltend man das am wenigsten nachteilige Spiel hat). Thus if $f(a)$ represents the probability of the value a being assumed by (fur) the unknown quantity x, it is not so important (ist weniger daran gelegen) that $f(a)$ should be a maximum as that $\int f(x) F(x-a)dx$, the integral extending over all possible values of x, should be a minimum ; when for F is selected a function that is continually positive and continually increases in a due degree (auf eine schichliche Art) with the increase of the variable. That the square is selected for this purpose is "purely arbitrary, and is in the nature of the subject that there should be this arbitrariness (Willkürlichkeit). Except for the well-known very great advantages....which the choice of the square secures, one might have chosen any other function satisfying the above conditions."

Rao has asked me why I thought it pertinent to consider the square error as a measure of the efficiency of the estimators which I studied. My reasons are the same as Gauss's.

that some minimum χ^2 estimators are more efficient [1] [2] [3] [4]. But he has presented no parallel analysis of these estimates, on the basis of some other loss function, disagreeing with this result. Until he does so, judgement of his criticism of my work should be held in abeyance.

Although at one point Rao criticises the use of the criterion of mean square error, at another point he seems to approve it, for he suggests that my "difficulties" are due to having applied it to the estimate of the wrong parameters. In my papers I was concerned with estimating parameters α and β (location and scale) of the logistic function. I shall explain that some 10 years ago, when I became concerned with the problem of statistical bio-assay, I found that not only were there innumerable articles but there were even several books concerned with this problem (or its equivalent in terms of other functions). It would seem that this itself is sufficient justification for examining the estimate of these parameters. But I may go further. Wishing to get data and examples in actual use, I communicated with several pharmaceutical firms, and from one very important house I received copious data of bio-assays that had been performed "by the method of Bliss" ("probits," with maximum likelihood). In all these assays a value of β was assumed as known from previous experience, and the problem was to estimate α (from which the E.D. 50 followed directly). This was the origin of my taking as the paradigmatic problem for mathematical statistical bio-assay the estimation of α with β known. Rao says I should instead have considered estimating the probabilities of death at various doses. But it is not for the mathematician to say what parameters should be estimated. It is his function only to say how parameters that are specified for him can best be estimated —if he can ! If Rao does not say that the P_i's should be estimated *instead of* α, β, but that it would be interesting and important to consider the estimate of the P_i's *also*, my *answer* is: "Yes, and I have thought of it, but with my limited and primitive means of computation it was important to do first things first, and besides, this particular programme is not so easy to define, much less to carry out, as it may appear."[9] If Rao desires it, I shall undertake some computations along these lines. I may say in advance, however, that (1) whatever the results may turn out to be, they will not mitigate the results already obtained in estimating α, β, which have their own primary importance and (2) I do not anticipate an essential reversal of my previous conclusion of the general relative inefficiency of the maximum likelihood estimates compared with some minimum χ^2 estimates, in these experiments.

We possess no principle of estimation the application of which ensures a best estimate in terms of the mean square error or any other objective operationally meaningful loss function. For the case of multinomial variation, I have defined an extended class of minimum χ^2 estimates which provides asymptotically efficient estimates [2], and this can frequently, but not always, be interpreted as estimates with approximately minimum variance in large samples. The maximum likelihood estimator can most simply be regarded as just one of the estimators in this class of minimum χ^2 estimators. For finite samples, really of any size but euphemistically referred to as small samples, we do not in general know which of these estimators has smallest mean square error. Certainly there is no reason to believe that the maximum likelihood is necessarily the best. The only way to find out is to investigate. Let us not stifle investigation by assuming that we already know. For some cases, I have found that what I have called the minimum transform χ^2 estimate has smaller mean square error than either the maximum likelihood estimator or the minimum Pearson estimator, and, incidentally, smaller than the lower bound for the variance of an unbiased regular estimator, which was widely thought to be impossible. For a special case with the logistic function I found another estimator—the Rao-Blackwellized estimator—which has even smaller mean square error. Mr. Joseph Hodges and I[6] will present, at the forthcoming Berkeley Symposium, another estimator, the H estimator, for the same case, which in a certain minimax sense is still better than the Rao-Blackwellized estimator. Different estimators can be developed for particular cases which have different operationally defined optimum properties. We do not have to have a monolithic statistics. Let investigation flower along different paths, and let a thousand estimators bloom!

I should like to take the opportunity to express my gratitude to Mr. Rao for his interest in my work and for the many invaluable suggestions that he has given me in the course of our correspondence. I have a very lively appreciation of the generosity reflected in so renowned a mathematician taking the trouble to help me. I realize that his presentation was not made to derogate my work, but to provoke a clarification of my views. I hope my attempt to reply in a forthright manner is understood in the same spirit.

---

[9] Which P_i's I should consider estimating ? The particular P_i's corresponding to the three experimental doses have no special interest.

REFERENCES

[1] BERKSON, J. (1957): Tables for use in estimating the normal distribution function by normit analysis. *Biometrika*, **44**, 411–435.

[2] ———— (1956): Estimating by least squares and by maximum likelihood. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1 University of California Press, Berkeley and Los Angeles, 1–11.

[3] ———— (1955): Estimate of the integrated normal curve by minimum normit chi-square with particular reference to bio-assay. *J.A.S.A.*, **50**, 529–549.

[4] ———— (1955): Maximum likelihood and minimum χ^2 estimates of the logistic function. *J.A.S.A.*, **50**, 130–162.

[5] BERKSON, J. and ELVEBACK, LILA (1960): Competing exponential risks, with particular reference to the study of smoking and lung cancer. *J.A.S.A.*, **55**, 415–428.

[6] BERKSON, J. and HODGES, J. L. JR. (1960): A minimax estimator for the logistic function. *Proceedings of the Fourth Berkeley Symposium on Probability and Statistics*. In press.

[7] EDGEWORTH, F. Y. (1908): On the probable errors of frequency-constants. *Jour. Roy. Stat. Soc.*, **71**, 386–87.

[8] FISHER, R. A. (1950): Theory of statistical estimation, in *Contributions to Mathematical Statistics*, John Wiley and Sons, Inc., New York, Paper 11, 700-725, See p. 714.

[9] MATHER, K. (1947): *Statistical Analysis in Biology*, Interscience Publishers, Inc., New York, Offset Lithoprint Reproduction, 212–213.

[10] MOOD, A. M. (1950): *Introduction to the Theory of Statistics*, McGraw Hill Book Company, Inc., New York, 160.

[11] QUENOUILLE, M. H. (1956): Notes on bias in estimation, *Biometetrika*. **43**, 353–360, 353.

[12] RAO, C. R. (1952): *Advanced Statistical Methods in Biometric Research*, John Wiley and Sons, Inc., New York, 139–140.

SIR RONALD FISHER: Mr. Neyman surprised many of us by his claim in his recent memorandum that Edgeworth introduced the Method of Maximum Likelihood. Edgeworth in fact bound his method on the theory of inverse probability and ascribed his notion specifically to K. Pearson and Filon in 1898; the Method of Maximum Likelihood may equally be found in this paper, only Pearson and Filon were under the misapprehension that the errors of random sampling were the same as those of the Method of Moments regarded as axiomatic by these authors.

Edgeworth, however, ends his paper with the reservation that all that he had said referred only to Measures of Central Tendency and not to the more complex problem of "The Fluctuation".

MR. KATZ lit les observations suivantes soumises par M.G.A. BARNARD.

It seems to the writer that the so-called anomalies in maximum likelihood estimation arises from misunderstanding of the problem which the method sets out to solve. The idea has grown up that the object of an estimation procedure is to find a single value for a parameter which may in some sense be regarded as "best", given a set of data. Alternatively, an interval is required within which the true value of the parameter may be supposed to lie. Neither of these formulations corresponds with the requirements of scientific inference. These can, in the first place, be roughly specified as requiring both a single value, to be regarded as "estimate" and indissolubly associated with it, some means of specifying the "error" to which this estimate is liable.

The method of maximum likelihood, in its simplest form, answers this requirement by giving as the "estimate" the point at which the log likelihood function has its maximum value, together with the inverse of the second derivative of this function at the maximum, which is used as an indication of the error. This procedure may be "justified" in several ways, but perhaps the principal justification can now be seen to consist in the facts: (1) that the log likelihood function is always minimal sufficient, so that for problems

of the type considered we need only aim to specify this function. (2) The log likelihood function is often well approximated in the neighbourhood of its maximum, by a quadratic expression; so that a specification of the location of the maximum, together with the second derivative there, gives us a good idea of the general course of the function.

From this point of view it is evident that we may expect "anomalies" to arise when the log likelihood function is far from being parabolic, and it is trivial that such instances can be constructed, starting from non-anomalous cases, by sufficiently pathological transformations of the parameters. More serious difficulties may arise when it is the form of the probability (density) function of the observations which makes the parabolic approximations poor —as may arise, for example, with certain configurations of small samples from the Cauchy distribution. In such cases we should bear in mind the principle of serendipity, according to which, if we are lucky enough to have obtained a sample which happens to give a parabolic log likelihood function, we need not concern ourselves with the problem of what we should have done had we been less lucky. In other cases, where serendipity does not come to our aid, we may either follow the suggestion made many years ago by Fisher, of specifying higher derivatives of the log likelihood, or we may use the sequence of moments of the likelihood function, rather than the sequence of its Taylor coefficients as the basis for our specification. In the case of the Cauchy distribution this would lead us to the Pitman estimator, though with an interpretation different from his, since we would think of it as associated with an "error" given by the second moment of the likelihood function, rather than as a "point estimate."

The problem of approximating to the specification of the log likelihood function, by way of the form indicated, is thus seen to have the same limited degree of arbitrariness associated with it as do other problems of approximation of functions.

In certain particular contexts a practical decision problem may be represented as leading to what has been called the problem of point estimation, and in such cases the loss function and a Bayesian prior distribution require to be specified before a unique solution can be arrived at. The fact that the data enter the solution of this problem through the likelihood function which they generate can be seen as another mode of justification of the likelihood approach. Evidently, under suitable regularity conditions, the solutions to wide classes of problems of this type could be seen to be estimators which are functions of the maximum likelihood estimator, together with the second and perhaps a few higher derivatives of the log likelihood function.

The need for a simplified description of the likelihood function by means of parabolic approximations, or otherwise, can be thought of as considerably reduced by the possibility, now existent, of drawing contours of constant likelihood, for up to 3 unknown parameters with the help of automatic computers. A specimen of such a contour map (for the programme for with I am indebted to Mr. H. Whitfield of Imperial College) for the unknown parameters p_1, p_2 arising from the 2×2 table is attached. The effect of the skewness of the likelihood function for p_2 can be seen quite clearly and the limitations of the paraboloidal approximation are apparent. It is also evident that these limitations would be reduced considerably if the logistic transformation

$$a_1 = \log p_1/(1-p_1), \quad a_2 = \log p_2/(1-p_2)$$

were applied to the parameters.

	A	not-A	Total	
	3	7	10	$P_r\{A\} = p_1$
	1	11	12	$P_r\{A\} = p_2$

All the essential ideas mentioned above seem to the present writer to have been implicit in Fisher's classical papers, and the only excuse for restating them here is that subsequent developments have shown that these classical papers have not always been studied with the attention they deserve.

APPARENT ANOMALIES AND IRREGULARITIES IN M. L. ESTIMATION

MR. BIRNBAUM: I would like to congratulate Mr. Rao on his presentation of very intresting contributions to the mathematical theory of estimation, and also to thank him for his clear statement of his general standpoint concerning the nature and purpose of the estimation problem. His view of estimation is a broad one, in which a single point estimate may be used in a variety of specific ways, some of them having the character of decision-making or specific inference problems, and some of them serving the purpose of efficient recording and interpretation of basic scientific or technical information of more general interest. Thus Mr. Rao's view of estimation is a kind of combination of the two standpoints presented by Messrs. Neyman and Barnard, and his general problem is that of showing how well a single point estimator can serve these broad and varied functions.

The standpoint which Mr. Neyman stated concisely is one upon which I based the first part of my own contribution here last week: in the problem of point-estimation, which formally includes confidence limit estimation, in general all possible estimators should be considered, and for a specified situation of application of choice of one estimator should (at least in principle) be used on comparisons of the probability distributions of all estimators. Such comparisons and choices may be informal or may utilize formal criteria, but they should reflect appropriately the situation of application and the statistician's purposes and judgements in the given situation—but the subject-matter of such comparisons and choices is basically those properties of estimators, represented by probabilities of errors of many kinds, which admit direct frequency interpretations. This standpoint leads in typical problems to large classes of admissible estimators, often including maximum likelihood estimators among many others. None of these admissible estimators can be eliminated from consideration as a matter of principle on the grounds mentioned; choices can be based only on grounds of specific judgements in specific problems and situations.

Mr. Barnard considers the point-estimation problem itself to be an incomplete and inadequate formulation of another inference problem. He states that the solution to this other problem is in principle the likelihood function itself, and that the role of the maximum likelihood point-estimate is simply to give a partial description of the likelihood function. What is this other problem whose solution is the likelihood function? I would call it the problem of informative inference, and define it as the problem of reporting efficiently, in meaningful objective terms, the statistical evidence, provided by an observed experimental outcome, which is relevant to the statistical hypotheses (possible parameter values) under consideration. Although the term "statistical evidence" is not in common use in mathematical statistics, I believe that it should be, because it represents accurately an essential feature of many important applications of statistical techniques. What is the nature of statistical evidence, and what are its objective qualitative and quantitative properties? As a familiar example, when an outcome of a scientific experiment indicates rejection of one statistical hypothesis in favour of another, on the basis of a test having very small probabilities of both types of errors, what seems most relevant and useful for typical purposes is the character of the outcome as strong evidence against the first hypothesis. It is a familiar fact that results of statistical tests are customarily interpreted in this way; one may wonder how often any standard statistical techniques would be used in scientific research if they did not admit such interpretations, which we may call evidential interpretations, of outcomes. The objective basis for interpreting the test outcome "reject" as strong evidence against a hypothesis is the small magnitude of its error-probabilities. For the familiar purpose of evidential interpretation of one given outcome of a test, it is enough that the latter probabilities admit an objective frequency interpretation in the conceptual sense. Certain relative frequencies which correspond to error-probabilities could in principle be realized physically, but will not be so realized in connection with the given experimental investigation; although this objective interpretation of these probabilities is purely conceptual, it suffices to support the interpretation of a single given outcome as statistical evidence.

The full analysis of the nature and structure of statistical evidence, in such objective probabilistic terms turns out to be a well-defined mathematical problem, as illustrated in the second part of my contribution here last week. Such analysis may be said to constitute a mathematical theory of informative statistical inference, and its subject-matter is quite distinct from intensities of belief or subjective probabilities. Such analysis leads to a certain central position for the likelihood function; and this analysis unfolds systematically, in objective probabilistic terms, the evidential significance inherent in the likelihood function itself. Such analysis gives support to the claim that informative inference should in principle

97

be based just on the likelihood function itself, and in my opinion eliminates the need for some of the other kinds of justification of this claim, mentioned by Mr. Barnard, which seem somewhat less direct; on the other hand, it seems essential to develop the general theory of informative inference, including complete explicit probabilistic interpretations of the statistical evidence provided by experiments of various mathematical forms. The simplest type of such interpretations is illustrated by the following example: when two simple hypotheses are considered, an outcome which gives the likelihood ratio statistic the value 99, regardless of the structure of the experiment in which it was obtained, has the same qualitative and quantitative properties, as evidence, as the outcome "reject" obtained by a statistical test having probabilities of errors of both kinds equal to .01. Such examples and interpretations illustrate the nature of the objective probabilistic bridge which can be constructed to connect systematically the two standpoints presented here by Mr. Neyman and Mr. Barnard.

I believe that such analysis clarifies certain essential unities and certain essential differences between the two standpoints mentioned, and that it can throw further light on the possibilities and possible limitations of programmes, such as that of Mr. Rao, which aim to go as far as possible in developing a single type of inference method which will prove satisfactory from both of these standpoints.

Mr. Rao : It is my first duty to express thanks to all those who contributed to the discussion. I have intentionally made some provocative statements in my paper to invite criticism necessary for a proper understanding of the issues involved. I think my plan has borne fruit. I would like to consider the various points raised in the discussion under a number of headings. The first one is historical.

1. HISTORICAL ASPECTS

I must admit I have not touched on the historical aspects of the m.l. method adequately, as that would be outside the scope of the subject assigned to me. But since Mr. Neyman raised some historical issues in his discussion I have to answer them.

Mr. Neyman states, "as far as I am aware, the priority in the approach of m.l.e..... belongs to F. Y. Edgeworth (1908a)"[10], a statement which Edgeworth himself would have contradicted as he attributed the method to Gauss, Laplace, and Pearson (footnotes on pages 384 and 395 of Edgeworth, 1908a). It also appears from Edgeworth's understanding of the earlier writers that the justification of m.l.e. consists in the inverse probability argument. Edgeworth (1908b, p. 500)[11] himself supported this view and was also aware of the contradictions involved in assigning the same a priori probability distributions to different functions of parameters, but contended that the matter was not serious in large samples and for functions not out of the ordinary (p. 392, Edgeworth, 1908a). It is, indeed, surprising that Mr. Neyman, paraphrasing Edgeworth's work, asserts that the estimate obtained by the method of inverse probability (i.e., by maximising the a posteriori distribution) is in fact the m.l. estimate. If $\hat{\theta}$ is an m.l. estimate of θ, then $\phi(\hat{\theta})$ is an m.l. estimate of any one-to-one function $\phi(\theta)$, while such a property is not true of estimates obtained by the inverse probability argument.

As for Laplace's work, it is clear from the interpretation by Todhunter (1865, p. 576, 585)[12] that Laplace never stressed the choice of "most probable result" nor did he justify its use in preference to any other method. I had no access to contributions by Gauss on this subject, but I take the liberty of quoting Mr. Barnard who thought that Gauss's justification of maximising the probability for estimation of parameters is not free from inverse probability.

[10] Edgeworth, F. Y. (1908a) : On the probable errors of frequency constants. *JRSS*, LXXI, 381.

[11] Edgeworth, F. Y. (1908b) : On the probable errors of frequency constants. *JRSS*, LXXI, 499.

[12] Todhunter, I. (1865) : *A History of the Mathematical Theory of Probability*. Chalsea Publishing Company, New York. (1949 edition).

APPARENT ANOMALIES AND IRREGULARITIES IN M. L. ESTIMATION

Karl Pearson (1896)[13] used the m.l. method in estimating the correlation coefficient without offering justification (p. 125), but his subsequent work with Filon (1898)[14] on standard errors of 'frequency constants' (estimates of parameters) did not show that he was considering m.l. estimates. It appears that they were attempting to compute the variance-covariance matrix of the asymptotic *a posteriori* distribution of the parameters presumably derived from a uniform *a priori* distribution and a sufficiently large sample. The analysis was not, however, rigorous.[15] We may thus infer that the authors were attempting to derive the asymptotic standard deviation of estimates which are the mean values of the *a posteriori* distribution, or those obtained by maximising the *a posteriori* probability density of the parameters. It is, however, somewhat puzzling to note that Pearson used the same expressions to determine the asymptotic standard errors of estimates obtained by the method of moments as well. It is also well known that Pearson did not advocate the use of m.l. in his subsequent writings.

A reference has also been made by Neyman to the result of Edgeworth, proved with the help of Love, that the most probable value (as determined by inverse probability) has the "smallest mean square deviation from the true point." This, indeed, is a remarkable attempt although the class of alternative estimates was very much restricted and the estimation was confined to location and scale parameters. A different argument is necessary to establish this result for the estimate of any general parameter and under less restrictive conditions on the class of estimates. The result, however, is not true, as observed by Hodges and mentioned by Neyman in the present discussion, without any restriction on the class of estimates.

We, therefore, do not have any literature supporting prior claims to the method of m.l.e., as a principle capable of wide application and justifying its use on reasonable criteria (such as efficiency in a sense wider than that used by Edgeworth and consistency) and not on inverse probability argument, before the fundamental contributions by Fisher in 1922 and 1925.

2. PHILOSOPHICAL STANDPOINT

Mr. Neyman wants me to explain my philosophical standpoint on estimation. "Should m.l.e. always be used irrespective of the properties it may have ?" If I understand correctly the spirit of this question and the emphasis on point estimation by Neyman (in the case of contamination of water) and by Berkson (in determining the concentration of a solution), I must differ from their philosophy quite sharply. I think for the estimation of contamination of water, instead of giving a point estimate, *sufficiently overestimated and considered safe* (in some sense), a statistician should ideally provide the customer with a whole series of inferences about the unknown value and the associated risks or consequences. For instance, in large samples under fairly general conditions, an estimate such as that obtained by m.l. together with its standard error estimable from the data themselves provides the complete answer. In small samples, mechanisms exist, under favourable circumstances, for providing fiducial probability statements or a whole series of fiducial limits in the sense of Fisher or confidence limits (interval, upper and lower) in the sense of Neyman.

I do not see how considerations of bias, under or over estimation arise. Again in the example of estimation of variance of an instrument, Neyman suggests that I should preferably give an unbiased estimate, if I have to escape "the lawsuit for damages by the manufacturer of the instrument." Assuming that a lawsuit is filed whenever there is an error in the estimate, an unbiased estimate can only give a mental consolation that errors made, however large they are, even out in the long run, although heavy damages may have to be paid every time ! It must be noted that if one adopts the "minimum mean square error" criterion for the choice of an estimate, the unbiased estimate may not even be admissible in the sense of decision theory. If the damage to be paid is proportional to the square of the error, I should not give an unbiased estimate.

[13] Pearson, K. (1896): Mathematical contributions to the theory of evolution IV. Regression, Heredity and Panmixia. *Phil. Trans. Roy. Soc.* London Series A, 187, 253.

[14] Pearson, K. and Filon, L. N. G. (1898): Mathematical contributions to the theory of evolution On the probable errors of frequency constants and on the influence of random selection on variation and correlation. *Phil. Trans. Roy. Soc.*, London, 191, 229.

[15]This problem is now under investigation and it is hoped to publish some of the results elsewhere.

I am aware that in some methodological problems such as obtaining a pooled estimate by averaging parallel estimates, one need to consider unbiased or nearly unbiased estimates. This can be achieved, in many cases, by a suitable adjustment of an available estimate.

What I maintain is that m.l.e. provides a convenient summary of data, demonstrably better than other methods in large samples, for answering questions of interest concerning an unknown parameter, and not that a point estimate obtained by maximising the likelihood is *the answer* to any specified question. Neyman and Berkson were repeatedly asking me during the discussion whether I would suggest the m.l. estimate in all situations. I do not know how the misunderstanding has arisen.

I am glad to note that Neyman looks upon "super efficient" estimates only as examples to show that the definition of efficiency as the attainment of minimum asymptotic variance is void without some restriction on the estimate. But I do not see why, *in large samples*, Hodges-LeCam "super efficient" estimates or the "super efficient" estimate given in the present paper for the mean of a normal population should not be preferred to $\bar{x}$, the sample mean (from behavioristic viewpoint). On the basis of decision theory, there is, perhaps, justification in doing so, or at least $\bar{x}$ has no definite claims over the other. My objection is, however, for other reasons. The super efficient estimate of the present paper is a function of the median and the mean of a sample of observations and is, therefore, less useful than $\bar{x}$, for purposes of statistical inference. The "super efficient" estimate of Hodges-LeCam is, however, equivalent to m.l.e. in large samples, i.e., efficient in the sense defined in the present paper and one may expect no substantial difference in the inferences associated with the two estimates in large samples. But it has certain defects. For instance, its asymptotic standard deviation being a discontinuous function of the unknown parameter does not admit reasonable estimation. Consequently, the inversion of a "super efficient" estimate for inference on the unknown parameter becomes a little complicated.

3. Minimum mean square error

It was not my intention to be unfair to Berkson in pointing out certain defects in the criterion of minimum mean square error. The example due to Silverstone of estimating the probability of success by the constant $1/2$ may be of a special nature. But we have a number of examples to illustrate that smaller variance does not necessarily mean higher concentration round the true value. It does not also imply that an estimate with a smaller variance provides a better discrimination between alternative values of the parameters. Recently, at my suggestion, Sethuraman (1960)[16] examined the relative powers of two statistics ξ and $2\xi + \eta$, (where ξ and η are the maximum and minimum respectively in a sample of size n from a rectangular population in the range $(\theta, 2\theta)$), for testing the hypothesis that $\theta = \theta_0$. Although as an estimate of θ, the m.l. estimate $\xi/2$ has uniformly larger variance than $(2\xi + \eta)/5$, an alternative estimate, it has better power as a test criterion for values of θ close to the assigned one. Since estimates with minimum mean square may not have *other* desirable properties, I was, naturally inclined to ask Berkson about the significance of, or the motivation for the choice of this criterion.

Berkson observes that if he follows my philosophy on theory of estimation, it will be disastrous in routine practice as in the use of a bio-assay for medical purposes, because one has to wait indefinitely collecting more and more observations before a decision can be reached. I have not said that decisions should not and cannot be made on the basis of available data, however meagre they are. But I am not convinced that an estimate which has minimum mean square error will be of help in minimising the mortality among his patients. I will only be too glad to accept Berkson's procedure if the latter were to be true. I am sure that for a statistical procedure to be made available for routine practice the approach should be somewhat different. Past data, as they accumulate, must be effectively used to improve the existing procedure. The theory of estimation as developed by Fisher is most suitable for such situations. It is not claimed anywhere that the m.l. estimates are minimal sufficient statistics, although they are explicit functions of the latter. It may be seen that in the problem of fitting a logistic function (specified by two parameters

---

[16] Sethuraman, J. (1960): Conflicting criteria of "Goodness" of statistics. *Sankhyā*, series A (in press).

α, β) to bio-assay data, the m.l. method does provide minimal sufficient statistics in samples where the m.l. estimates are properly defined. If a suitable convention of specifying an estimate, like the one suggested by Silverstone, is adopted in the case of samples for which the m.l. estimates of one or both of α and β are not finite, sufficiency of m.l. estimates can be claimed for all samples.[17] The situation is not so simple in the case of minimum logit χ^2, advocated by Berkson. Generally, the estimates obtained by this method are not sufficient. But Berkson insists on quoting one example with 3 doses and 10 animals at each dose, in which case, with the application of a dubious rule such as $2n$-th, the minimum logit χ^2 estimate happens to be sufficient.

Berkson gives no other argument in favour of mean square error, except that "it is a representative loss function, and if, in a particular application, some other loss function, suggests itself, let it be investigated : suppose $\hat{\alpha}$ and α^* are two alternative estimates of a parameter α such that,

$$E(\hat{\alpha}-\alpha)^2 > E(\alpha^*-\alpha)^2$$

and there exists a function ϕ such that

$$E_n[\phi(\hat{\alpha})-\phi(\alpha)]^2 < E[\phi(\alpha^*)-\phi(\alpha)]^2$$

then the estimate of α obtained by using one loss function is not good with respect to another loss function. Such situations are not rare and any number of examples with a reasonable choice of the function ϕ can be given. In the problem of fitting the logistic function, I venture to suggest that some increasing function of the differences between hypothetical and estimated probabilities of success or failure at each dose, may be a better indicator of the goodness of estimation than the deviations in the estimates of parameters α and β themselves. I do not know whether a minimum logit χ^2 estimate would have smaller expected loss than other types of estimates when loss functions of the type indicated are considered.

4. OTHER ASPECTS

The views expressed by Barnard on point estimation, the role it plays in specifying the likelhoood, and its relation to a practical decision problem do not seem to be in conflict with those in my paper. Both of us have tried to interpret Fisher's work on estimation, though not completely and not in exactly the same way. I hope they will serve to remove some wrong notions about m.l. found in recent literature.

I am particularly interested in Birnbaum's contribution to the theory of estimation as it provides a small sample justification to certain estimation procedures including the m.l. This is a far more difficult task than what I have attempted to do confining my remarks mainly to the case of large samples. I cannot think of situations where serious decisions are taken on meagre evidence supplied by small samples, while in routine practice such as the application of control charts in industry one may think of specifying rules of action based even on very small samples to minimise certain risks in the long run. Further discussion on the theory of estimation in small samples as attempted by Birnbaum would, no doubt, be of great value.

I would also wish to take the opportunity of mentioning a few results in connection with the investigation mentioned in the last paragraph of my paper. It was thought that no distinction could be made in large samples among estimation procedures such as m.l., minimum chi-square, modified minimum chi-square, etc. since they all provide asymptotically efficient estimates in a wider sense of $i_T \rightarrow i$ (i_T and i are informations, *per observation* contained in the statistic T and the sample respectively). But as mentioned by Fisher in the 1925 paper, differences in the *actual* amounts of information contained in different estimates are more relevant. It has been possible to compute a quantity, analogous to, if not same as, the limiting difference in the total information contained in the statistic and in the sample and establish that the m.l. method has the least limiting loss. The minimum chi-square, modified minimum chi-square, and other related methods have a greater loss. The actual values are given by the author in a paper under print in the Proceedings of the 4th Berkeley Symposium on Statistics and Probability.

---

[17] The emphasis should be not on estimating the parameters α and β but on probabilities of death at various doses. The parameter space has then to be properly defined in terms of these probabilities. Once this is done, many difficulties mentioned by Dr. Berkson would disappear.

ACKNOWLEDGEMENT

CONTENTS

DENSITY IN THE LIGHT OF PROBABILITY THEORY

By E. M. PAUL

Indian Statistical Institute

SUMMARY. Let $p_1 = 2, p_2, p_3, ...,$ be the prime numbers in ascending order. Let $\{X_n\}$ be a sequence of measure spaces, each X_n consisting of the points $0, 1, 2, 3, ...$. In X_n, we place mass $\left(1 - \dfrac{1}{p_n}\right) \cdot \dfrac{1}{p_n^r}$ at the point $r(r = 0, 1, 2, ...)$. We take the product space $X_1 X_2 X_3 ...$ and the product measure P in it. Each point of this space is an 'infinite vector' $(x_1, x_2, ...)$, the coordinates being nonnegative integers.

Now let S be any set of positive integers. By the *upper magnification* $M^U(S)$ of S, we mean the set of vectors $(x_1, x_2, ...)$ such that $2^{x_1} p_2^{x_2} ... p_n^{x_n} \epsilon S$ for infinitely many values of n. By the *lower magnification* $M_L(S)$ of S, we mean the set of vectors $(x_1, x_2, ...)$ such that $2^{x_1} p_2^{x_2} ... p_n^{x_n} \epsilon S$ for all sufficiently large n. In this paper, we prove that $P[M_L(S)] \leqslant \delta_L(S) \leqslant \delta^U(S) \leqslant P[M^U(S)]$, where $\delta_L(S)$ and $\delta^U(S)$ represent the lower and upper logarithmic densities of S, respectively.

Also, let f be a real-valued function defined on the set of positive integers. We prove that if the sequence of random variables $f\left(2^{x_1} p_2^{x_2} ... p_n^{x_n}\right)$, defined on the probability space $X_1 X_2 X_3 ...$ converges with probability one to a random variable g, then f has a distribution, namely, that of g; in defining the distribution of f, we employ logarithmic density.

In this paper, we formulate the whole theory in an abstract framework.

1. INTRODUCTION

The concept of density of a set of positive integers has close connections with the general idea of probability. Still, when one tries to employ the machinery of the modern theory of probability in the investigation of density, many difficulties are encountered. The basic reason for this situation stems from the fact that density is not, in general, countably additive. In this paper, we study density by embedding the set of positive (or nonnegative) integers in a suitable probability space. In order to be able to tackle different kinds of problems, we develop the theory axiomatically and demonstrate some concrete situations to which the axiomatic theory is applicable.

2. THE SPACE X

We consider a sequence (X_n) of abstract spaces. Each X_n will consist of a sequence of points $p_n^0, p_n^1, p_n^2, \ldots$. We form the product space $\overset{\infty}{\underset{n=1}{X}} (X_n) = X$. Thus an element of X will be a 'vector' $(p_1^{\alpha_1}, p_2^{\alpha_2}, \ldots)$ where each α_r is a nonnegative integer. We shall also denote this element by $p_1^{\alpha_1} p_2^{\alpha_2} \ldots$. However, we do not have any algebraical operation in view. We may also denote the element $p_1^{\alpha_1} p_2^{\alpha_2} \ldots$ by $(\alpha_1, \alpha_2, \ldots)$. I will stand for the set of vectors (in X) having only a finite number of positive coordinates. For $1 \leqslant r_1 < r_2 < r_3 \ldots < r_k$ and $m_1, m_2, \ldots, m_k$ all $\geqslant 0$, *by the set of vectors in I corresponding to* $(p_{r_1}^{m_1}, p_{r_2}^{m_2}, \ldots, p_{r_k}^{m_k})$ we shall mean the set of those vectors in I whose r_1-th coordinate is $m_1, \ldots, r_k$-th coordinate is m_k. We shall denote this set by $S(p_{r_1}^{m_1}, \ldots, p_{r_k}^{m_k})$. We shall often denote the general vector $(x_1, x_2, x_3, \ldots)$ by x.

We now associate with every subset σ of I a real number $\delta^U(\sigma)$ satisfying Postulates (A) to (F), σ' will stand for $(I-\sigma)$.

(A) $0 \leqslant \delta^U(\sigma) \leqslant 1$ for every σ.

(B) If σ_1 and σ_2 are two subsets of I and $\sigma_1 \subset \sigma_2$, $\delta^U(\sigma_1) \leqslant \delta^U(\sigma_2)$.

(C) If σ_1 and σ_2 are two disjoint sets,

$$\delta^U(\sigma_1 \cup \sigma_2) \leqslant \delta^U(\sigma_1) + \delta^U(\sigma_2)$$

and

$$\delta^U(\sigma_1') + \delta^U(\sigma_2') \geqslant 1 + \delta^U(\sigma_1' \cap \sigma_2').$$

(D)

$$\delta^U(I) = 1.$$

$$\delta^U(\phi) = 0, \phi \text{ being the empty set } (< I).$$

Before introducing Postulates (E) and (F), we frame

Definition 1 : For every $\sigma(\subseteq I)$, we define $\delta_L(\sigma)$ to be $= 1 - \delta^U(\sigma')$.

Proposition 1 : For every $\sigma, 0 \leqslant \delta_L(\sigma) \leqslant \delta^U(\sigma) \leqslant 1$.

By Postulates (C) and (D),

$$1 = \delta^U(I) = \delta^U(\sigma \cup \sigma') \leqslant \delta^U(\sigma) + \delta^U(\sigma').$$

$$\therefore \quad \delta_L(\sigma) = 1 - \delta^U(\sigma') \leqslant \delta^U(\sigma).$$

Since

$$\delta^U(\sigma') \leqslant 1, \quad 1 - \delta^U(\sigma') \geqslant 0.$$

$$\therefore \quad \delta_L(\sigma) \geqslant 0.$$

We note that $\delta_L(\phi) = 1 - \delta^U(\phi') = 1 - \delta^U(I) = 0$; $\delta_L(I) = 1 - \delta^U(\phi) = 1$.

Proposition 2 : If $\sigma_1 \subseteq \sigma_2$, $\delta_L(\sigma_1) \geqslant \delta_L(\sigma_2)$. Since $\sigma_1' \subseteq \sigma_2'$, $\delta^U(\sigma_1') \leqslant \delta^U(\sigma_2')$, by Postulate (B).

$$\therefore \quad 1 - \delta^U(\sigma_1') \geqslant 1 - \delta^U(\sigma_2').$$

$$\therefore \quad \delta_L(\sigma_1) \geqslant \delta_L(\sigma_2).$$

Definition 2 : If σ is such that $\delta_L(\sigma) = \delta^U(\sigma)$, we denote the common value by $\delta(\sigma)$.

Proposition 3 : δ is finitely additive; that is, if σ_1 and σ_2 are two disjoint subsets of I such that $\delta(\sigma_1)$ and $\delta(\sigma_2)$ exist, then $\delta(\sigma_1 \cup \sigma_2)$ exists and $= \delta(\sigma_1) + \delta(\sigma_2)$.

Proof: By Postulate (C), $\delta^U(\sigma_1 \cup \sigma_2) \leqslant \delta^U(\sigma_1) + \delta^U(\sigma_2)$, and the latter $= \delta_L(\sigma_1) + \delta_L(\sigma_2) = 1 - \delta^U(\sigma_1') + 1 - \delta^U(\sigma_2') \leqslant 2 - (1 + \delta_U(\sigma_1' \cap \sigma_2')) = 1 - \delta^U(\sigma_1' \cap \sigma_2')$ $= \delta_L(\sigma_1 \cup \sigma_2)$.

$$\therefore \quad \delta^U(\sigma_1 \cup \sigma_2) = \delta_L(\sigma_1 \cup \sigma_2) = \delta(\sigma_1) + \delta(\sigma_2).$$

Proposition 4 : If (σ_n) is a sequence subsets of disjoint of I and each σ_n has a δ,
$$\delta_L(\bigcup_n \sigma_n) \geqslant \sum_n \delta(\sigma_n).$$

Proof: By Proposition 3, $\bigcup_{n=1}^{m} \sigma_n$ has $\delta = \sum_{n=1}^{m} \delta(\sigma_n)$, for every m.

Now
$$\bigcup_n \sigma_n \supseteq \bigcup_{n=1}^{m} \sigma_n.$$

So by Proposition 2,
$$\delta_L(\bigcup_n \sigma_n) \geqslant \delta_L(\bigcup_{n=1}^{m} \sigma_n) = \delta(\bigcup_{n=1}^{m} \sigma_n) = \sum_{n=1}^{m} \delta(\sigma_n).$$

Since this is true for all m, $\delta_L(\bigcup_n \sigma_n) \geqslant \sum_{n=1}^{\infty} \delta(\sigma_n)$.

Proposition 5 : If $\sigma (\subset I)$ is such that $\delta^U(\sigma) + \delta^U(\sigma') = 1$, then $\delta(\sigma)$ exists.

Proof:
$$\delta_L(\sigma) + \delta_L(\sigma') = 1 - \delta^U(\sigma') + 1 - \delta^U(\sigma) = 1.$$
So
$$\delta^U(\sigma) + \delta^U(\sigma') = 1, \; \delta_L(\sigma) + \delta_L(\sigma') = 1.$$
So by Proposition 1,
$$\delta^U(\sigma) = \delta_L(\sigma), \; \delta^U(\sigma') = \delta_L(\sigma').$$
So $\delta(\sigma)$ and $\delta(\sigma')$ exist.

Corollary : *If σ is such that $\delta_L(\sigma) + \delta_L(\sigma') = 1$, then $\delta(\sigma)$ exists.*
$$1 - \delta^U(\sigma') + 1 - \delta^U(\sigma) = 1$$
$$\therefore \quad \delta^U(\sigma) + \delta^U(\sigma') = 1.$$

Postulate (E) : For $1 \leqslant r_1 < r_2 < \ldots < r_k$ and $m_1, \ldots, m_k$ all $\geqslant 0$,
$$\sigma = S\left(p_{r_1}^{m_1}, \ldots, p_{r_k}^{m_k} \right) \text{ has } \delta_L(\sigma) = \delta^U(\sigma) = \delta(\sigma). \text{ We shall denote the } \sigma \text{ associated}$$
with $S\left(p_1^{\alpha_1}, \ldots, p_n^{\alpha_n} \right)$ by $\delta_{\alpha_1, \ldots, \alpha_n}$.

Postulate (F) : For $m \geqslant 1$, $\Sigma \delta(\sigma) = 1$ where $\sigma = S(p_1^{n_1}, p_2^{n_2}, \ldots, p_m^{n_m})$, the summation extending over all vectors $(n_1, n_2, \ldots, n_m)$ with nonnegative integral coordinates.

Proposition 6 : Let $0 \leqslant \beta_1 < \beta_2 < \ldots$ be a finite or infinite sequence of integers. Then $\bigcup_m S(p_1^{\beta_m})$ has a $\delta = \sum_m \delta[S(p_1^{\beta_m})]$.

Proof: If there are only finitely many β's, the result is immediate since δ is finitely additive (Proposition 3). So suppose there are infinitely many β's. Let $\Gamma = (\gamma_1 < \gamma_2 < \ldots)$ be the set of nonnegative integers complementary to the set B of β's.

By Proposition 4,

$$\delta_L[\bigcup_m s(p_1^{\beta m})] \geqslant \sum_m \delta[S(p_1^{\beta m})]. \qquad \ldots \text{(2.1)}$$

Similarly,

$$\delta_L[\bigcup_m S(p_1^{\gamma n})] \geqslant \sum_n \delta[S(p_1^{\gamma n})]. \qquad \ldots \text{(2.2)}$$

Now by Postulate (F), $\quad \sum_m \delta[S(p_1^{\beta m})] + \sum_n \delta[S(p_1^{\gamma n})] = 1. \qquad \ldots \text{(2.3)}$

From (2.1), (2.2) and (2.3) we get

$$\delta_L[\bigcup_m S(p_1^{\beta m}) = A] + \delta_L[\bigcup_m S(p_1^{\gamma n}) = B] \geqslant 1.$$

$$1 - \delta^U(A) + 1 - \delta^U(B) \geqslant 1.$$

$$\therefore \qquad \delta^U(A) + \delta^U(B) \leqslant 1.$$

Now A and B are complementary sets. So by Postulate (C),

$$\delta^U(A) + \delta^U(B) \geqslant \delta^U(I) = 1.$$

So $\qquad \delta^U(A) + \delta^U(B) = 1. \qquad \ldots \text{(2.4)}$

So by Proposition 5, $\delta(A)$ and $\delta(B)$ exist. Now we write (2.1) and (2.2) as

$$\delta(A) \geqslant x,$$
$$\delta(B) \geqslant y.$$

(2.3) becomes $\qquad x + y = 1$

(2.4) becomes $\qquad \delta(A) + \delta(B) = 1.$

$$\therefore \qquad \delta(A) = x, \; \delta(B) = y.$$

$$\therefore \qquad \delta[\bigcup_m S(p_1^{\beta m})] = \sum_m \delta[S(p_1^{\beta m})].$$

On the basis of the foregoing six postulates we introduce a probability distribution in space X. In the space X_1, we place at $p_1^{n_1}$ mass $= \delta[S(p_1^{n_1})]$, $n_1 = 0, 1, 2, 3, \ldots$. By Postulate (F), the total mass is 1. Similarly in the space $X_1 X_2$, we place at $(p_1^{n_1}, p_2^{n_2})$ mass $= \delta[S(p_1^{n_1}, p_2^{n_2})]$,

$$\begin{matrix} n_1 \\ n_2 \end{matrix} = 0, 1, 2, 3, \ldots$$

In this way we get a probability distribution in $X_1 X_2 \ldots X_k$-space for $k = 1, 2, 3, \ldots$. These distributions are mutually consistent. For example, we shall prove that the distribution in $X_1 X_2$ is consistent with that in X_1.

$$S(p_1^{m_1}) = S(p_1^{m_1}, p_2^0) \bigcup S(p_1^{m_1}, p_2') \bigcup \ldots$$

By (B), and by Proposition 3,

$$\delta_{m_1} \geqslant \delta_{m_1,0} + \delta_{m_1,1} + \ldots \text{ to } n \text{ terms.}$$

$$\therefore \quad \delta_{m_1} \geqslant \delta_{m_1,0} + \delta_{m_1,1} + \ldots \text{ ad inf.}$$

This is true for $m_1 = 0, 1, 2, 3, \ldots$ So adding these inequalities,

$$\overset{\infty}{\underset{m_1=0}{\Sigma}} \delta_{m_1} \geqslant \overset{\infty}{\underset{m_1=0}{\Sigma}} \left\{ \overset{\infty}{\underset{n_1=0}{\Sigma}} \delta_{m_1,n_1} \right\} = \overset{\infty}{\underset{m_1=0}{\Sigma}} \overset{\infty}{\underset{n_1=0}{\Sigma}} \delta_{m_1,n_1} .$$

Now by Postulate (F), each extreme member has value 1. So $\delta_{m_1} = \delta_{m_1, 0} + \delta_{m_1, 1} + \cdots$ ad inf for each $m_1 = 0, 1, 2, \ldots$. In this way, we see that the distributions in $X_1, X_1 X_2, X_1 X_2 X_3, \ldots$ are all consistent. So by Kolmogorov's Theorem (1956, p. 27), we get a unique probability distribution in X.

We shall denote this measure by P.

Proposition 7 : Let $1 \leqslant r_1 < r_2 < \ldots < r_k$. Let $\left(p_{r_1}^{m_{n_1}}, \ldots, p_{r_k}^{m_{n_k}} \right), n = 1, 2, \ldots$ be a finite or infinite set of points in the $X_{r_1} X_{r_2} \ldots X_{r_k}$-space. Then

$$\delta \left\{ \bigcup_n S \left(p_{r_1}^{m_{n_1}}, \ldots, p_{r_k}^{m_{n_k}} \right) \right\} \text{ exists}$$

and

$$= \underset{n}{\Sigma} \, \delta \left\{ S \left(p_{r_1}^{m_{n_1}}, \ldots, p_{r_k}^{m_{n_k}} \right) \right\} = \underset{n}{\Sigma} \, P \left\{ C \left(p_{r_1}^{m_{n_1}}, \ldots, p_{r_k}^{m_{n_k}} \right) \right\}$$

where $C \left(p_{r_1}^{m_{n_1}}, \ldots, p_{r_k}^{m_{n_k}} \right)$ denotes the cylinder set formed by vectors having m_{n_1} as the r_1-th coordinate, $\ldots$ and m_{n_k} as the r_k-th coordinate.

Proof: We observe that $C \left(p_{r_1}^{m_{n_1}}, \ldots, p_{r_k}^{m_{n_k}} \right)$ may be looked upon as the union of countably many disjoint cylinder sets, each of these sets having as base a single point in $X_1 X_2 X_3 \ldots X_{r_k}$-space. The rest of the proof is exactly parallel to that of Proposition 6.

3. THE MAGNIFICATION THEOREM

Let S be any subset of I. The set M_U of vectors $(x_1, x_2, \ldots) \epsilon X$ such that $(x_1, x_2, \ldots, x_k, 0, 0, \cdots) \epsilon S$ for infinitely many values of k will be called the *upper magnification* of S. The set M_L of vectors $(x_1, x_2, \ldots) \epsilon X$ such that $(x_1, \ldots, x_k, 0, 0, \cdots) \epsilon S$ for all sufficiently large values of k will be called the *lower magnification* of S. Clearly, $M \subseteq M_U$ and both sets are measurable.

We also employ the notations $M_U(S)$ and $M_L(S)$. If $M_L(S) = M_U(S)$, we shall denote it by $M(S)$.

Let $S' = I - S$. Then $M_L(S') = X - M_U(S)$ and $M_U(S') = X - M_L(S)$.

We now introduce two conditions. Let I_k be the subset of I consisting of vectors having nonzero coordinates only in the first k places (at most).

Condition G: $\delta(I_k) = 0$, $k = 1, 2, 3, \ldots$. We shall call a subset S of I *right complete* in case $(x_1, x_2, \ldots, x_n, 0, 0, \cdots) \epsilon S$ implies $(x_1, \ldots, x_n, x_{n+1}, \ldots, x_{n+m}, 0, 0, \ldots) \epsilon S$ for every $m \geqslant 0$ and arbitrary nonnegative integers $x_{n+1}, \ldots, x_{n+m}$. It is easy to verify that if S is right-complete, $M_{\bar{U}}(S) = M_L(S)$. We note that a right-complete set containing $(0, 0, 0, \ldots)$ must be the whole of I.

Condition H : Every right-complete set T is such that $\delta(T)$ exists and $\delta(T) = P[M(T)]$.

We shall formulate condition H in a slightly different form. We shall call a subset T of I *left-complete* in case $(x_1, x_2, \ldots, x_n, 0, 0, \ldots) \epsilon T$ implies $(x_1, \ldots, x_{n-1}, 0, 0, 0, \ldots) \epsilon T$ for $n = 1, 2, 3, \ldots$. Thus every nonempty left-complete set contains the vector $(0, 0, 0, \ldots)$. If T is left-complete, $M_U(T) = M_L(T)$. If T is left-complete, $(I-T)$ is right-complete and vice versa. Condition H is equivalent to condition H_1.

Condition H_1 : Every left-complete set T is such that $\delta(T)$ exists and $\delta(T) = P[M(T)]$.

Theorem 1 : *If conditions G and H hold,*

$$P[M_U(S)] \geqslant \delta^U(S)$$

where S is an arbitrary subset of I.

Proof: For $k = 1, 2, 3, \ldots$, let G_k be the set of vectors $(x_1, x_2, x_3, \ldots)$ such that $(x_1, \ldots, x_k, 0, 0, \ldots)$, $(x_1, \ldots, x_{k+1}, 0, 0, \ldots)$, $(x_1, \ldots, x_{k+2}, 0, 0, 0, \ldots)$, $\ldots$ all $\epsilon S' = I - S$. We easily verify that $G_n \leqslant G_{n+1}$ for $n = 1, 2, 3, \ldots$ and that $\bigcup_1^\infty G_n = X - M_U(S)$. Take any $\epsilon > 0$. Let $k(\epsilon) = k$ be such that $P[G_k] > P[X - M_U(S)] - \epsilon$. Consider the set B_ϵ of vectors $(0, 0, 0, \ldots)$, $(x_1, 0, 0, \ldots)$, $(x_1, x_2, 0, 0, \ldots)$, $\ldots$ where the vector $(x_1, x_2, x_3, \ldots)$ runs through $G_{k(\epsilon)}$.

B_ϵ is a left-complete set. So $\delta(B_\epsilon)$ exists and is $= P[M_L(B_\epsilon)] \geqslant P[G_k(\epsilon)] > P[X - M_U(S)] - \epsilon$. Now let D_ϵ be the set of vectors $(x_1, \ldots, x_{k(\epsilon)}, 0, 0, \ldots)$, $(x_1, \ldots, x_{k(\epsilon)+1}, 0, 0, \ldots)$, $(x_1, \ldots, x_{k(\epsilon)+2}, 0, 0, \ldots)$, $\ldots$ where $(x_1, x_2, x_3, \ldots)$ runs through $G_{k(\epsilon)}$. Then $\delta_L(D_\epsilon^-) = \delta_L(B_\epsilon)$ by condition G and Postulate (C). So $\delta_L(S') \geqslant \delta_L(D_\epsilon) = \delta_L(B_\epsilon) = \delta(B_\epsilon) > P[X - M_U(S)] - \epsilon$. Since $\epsilon > 0$ is arbitrary,

$$\delta_L(S') \geqslant P[X - M_U(S)].$$

$$\therefore \quad \delta^U(S) = 1 - \delta_L(S') \leqslant P[M_U(S)].$$

Corollary (The Magnification Theorem) : *If conditions G and H hold, then for every subset S of I.*

$$P[M_L(S)] \leqslant \delta_L(S) \leqslant \delta^U(S) \leqslant P[M^U(S)].$$

Proof: We apply Theorem 1 to $S' = 1 - S$ and get $P[M^U(S')] \geqslant \delta^U(S')$. But $M^U(S') = X - M_L(S)$ and $\delta_L(S) = 1 - \delta^U(S')$.

In view of the importance of condition H, we shall examine the structure of right-complete sets. Let S be a right-complete set. If $(0, 0, ...) \epsilon S$, $S = I$. So let us suppose $(0, 0, 0, ...) \notin S$. We shall call $(x_1, ..., x_n, 0, 0, 0, ...)$ a *basic vector* of (or in) S in case $(x_1, ..., x_n, 0, 0, ...) \epsilon S$ but $(x_1, ..., x_{n-1}, 0, 0, ...) \notin S$; if $n = 1$, we interpret $(x_1, ..., x_{n-1}, 0, 0, ...)$ as $(0, 0, 0, ...)$. If $(x_1, ..., x_n, 0, 0, ...)$ is a basic vector, $x_n > 0$. If $v \epsilon S$, there is a unique basic vector $\zeta = (x_1, ..., x_n, 0, 0, 0, ...)$, $n \geqslant 1$, such that $v = (x_1, ..., x_n, y_{n+1}, ..., y_{n+m}, 0, 0, 0, ...)$ for some $m \geqslant 0$. Conversely, if $(x_1, ..., x_n, 0, 0, ...)$ is a basic vector of S, $(s_1, ..., x_n, y_{n+1}, ..., y_{n+m}, 0, 0, 0, ...) \epsilon S$, where $m, y_{n+1} ..., y_{n+m}$ are any nonnegative integers. Thus the basic vectors of S completely determine S. Let the basic vectors of S be $\zeta_1, \zeta_2, ...$ where $\zeta_n = (x_1, ..., x_{l_n}, 0, 0, ...)$, $x_{r_n} \geqslant 1$. Then $M(S) = \bigcup_n C_n$ where C_n is the cylinder set formed by vectors whose first r_n coordinates are $x_1, ..., x_{r_n}$ in that order; $C_i \bigcap C_j$ is empty if $i \neq j$.

4. DISTRIBUTIONS OF ARITHMETICAL FUNCTIONS

We refer to Section 2. We have the space X and the probability measure P in it.

Let f be a finite real-valued function defined on the subset I of X. f *will be said to have the distribution* Q in case Q is a probability distribution on $(-\infty, \infty)$ and for every c such that $Q(c) = 0$

$$E_n\{n \epsilon I \text{ and } f(n) < c\} \text{ has } \delta = Q(-\infty, c).$$

If f has a distribution, it is unique, since two distributions are identical in case they assign the same measure to $(-\infty, c)$ for every common continuity point c.

Suppose f is a real finite-valued function defined in the set I.

Theorem 2: *Suppose* $g(x) = \lim_{m \to \infty} f(p_1^{x_1} p_2^{x_2} ... p_m^{x_m})$ *exists at almost all points* $x = (x_1, x_2, ...)$ *of* X.

Moreover, let δ^U *satisfy conditions* G *and* H *of Section* 3. *Then* f *has a distribution and this is the same as the distribution of* $g(x)$.

Proof: Let Q be the distribution of $g(x)$. Let c be such that $Q(c) = 0$. Let E_c be the set of points $x \epsilon X$ such that $g(x) < c$. Let S_c be the set of vectors $n \epsilon I$ such that $f(n) < c$. Suppose $x = (x_1, x_2, x_3, ...) \epsilon E_c$. Then $\lim_{m \to \infty} f(p_1^{x_1} p_2^{x_2} ... p_m^{x_m}) < c$; so for all sufficiently large values of m, $f(p_1^{x_1}, ... p_m^{x_m}) < c$. So for all sufficiently large values of m, $p_1^{x_1} p_2^{x_2} ... p_m^{x_m} \epsilon S_c$. Thus $(x_1, x_2, ...) \epsilon M_L(S_c)$. So $E_c \subseteq M_L(S_c)$.

$$\therefore \quad \delta_L(S_c) \geqslant P[M_L(S_c)] \geqslant P(E_c) = Q(-\infty, c).$$

Similarly if T_c is the set of vectors $n \epsilon I$ such that $f(n) > c$, $\delta_L(T_c) \geqslant Q(c, \infty)$.

$$\therefore \quad \delta^U(S_c) = 1 - \delta_L(S_c') \leqslant 1 - \delta_L(T_c)$$

$$\leqslant 1 - Q(c, \infty) = Q(-\infty, c), \text{ since } Q(c) = 0.$$

But we have $\qquad \delta_L(S_c) \geqslant Q(-\infty, c).$ $\qquad \therefore \quad \delta(S_c) = Q(-\infty, c).$

5. Integration with respect to density

Suppose f is a bounded real-valued function defined on the set $I(\subseteq X)$. Let $A_1, ..., A_n$ be disjoint subsets of I such that $\bigcup_1 A_i = I$ and such that $\delta(A_i)$ exists for $i = 1, ..., n$; let us call this partition P. Let

$$S(P) = \sum_{i=1}^n \delta(A_i) \sup_{x \, \varepsilon \, A_i} f(x),$$

$$s(P) = \sum_{i=1}^n \delta(A_i) \inf_{x \, \varepsilon \, A_i} f(x).$$

As usual, we call the lower bound of $S(P)$ the upper integral of f with respect to δ and denote it by $\overline{\int} fd(\delta)$. Similarly the lower integral $\underline{\int} fd(\delta)$.

Theorem 3 : *Let f be any bounded real-valued function defined on I.*

For every right-complete subset S of I, let $M(S)$ be such that $P[M(S)] = \delta^U(S)$ $= \delta_L(S)$. Let for each $x = (x_1, x_2, ...) \, \varepsilon \, X$, $g(x) = \underset{n \geqslant 1}{L.U.B.} \ f(x_1, ..., x_n, 0, 0, ...)$. Then

$$\overline{\int} fd(\delta) \leqslant \int_X g(x)dP.$$

Proof : Suppose $-K < g < +K$. Take any $\epsilon > 0$.

Partition $(-K, +K)$ into a finite number of sub-intervals in such a way that (i) the end-points of the sub-intervals are all continuity points in the distribution of the random variable $g(x)$, and (ii) every approximative sum for $\int_X g(x)dP$ formed on the basis of this partition lies between $\int_X gdP - \epsilon$ and $\int_X gdP + \epsilon$. Let (y_k, y_{k+1}) be a typical sub-interval in this partition; let A_k be the subset of X on which $g(x) > y_k$. $A_k \cap I$ is a right-complete set; and A_k is the magnification of this set. Hence

$$\delta\{(A_k - A_{k+1}) \cap I\} = P(A_k - A_{k+1}) = Pr\{y_k < g(x) < y_{k+1}\}.$$

Now partition I into the subsets $(A_k - A_{k+1}) \cap I$. Upper approximate sum for $\int fd(\delta)$ given by this partition is $\leqslant$

$$\sum_k y_{k+1}\delta\{(A_k - A_{k+1}) \cap I\} = \sum_k y_{k+1} \ Pr\{y_k < g(x) < y_{k+1}\}.$$

Now the latter expression is $\qquad < \int_X g(x)dP + \epsilon.$

Thus $\qquad \int fd(\delta) \leqslant \int_X g(x)dP.$

In the foregoing theorem, $g(x)$ was defined as $\underset{n \geqslant 1}{L.U.B.} \ f(x_1, ..., x_n, 0, 0, ...)$. We shall show that if condition G also holds, the proof will hold if $g(x)$ is defined as $\underset{n \geqslant m}{L.U.B.} \ f(x_1, ..., x_n, 0, 0, ...)$ where m is any positive integer. Taking the limit as

$m \to \infty$, we get the

Corollary : *Under conditions G and H,*

$$\overline{\int} f d(\delta) \leqslant \int_X \{\lim_n \sup f(x_1, ..., x_n, 0, 0, ...)\} dP.$$

Similarly
$$\underline{\int} f d(\delta) \geqslant \int_X \{\lim_n \inf f(x_1, ..., x_n, 0, 0, ...)\} dP.$$

We now give the proof for the case where $g(x)$ is defined as L.U.B. $f(x_1, ..., x_n,$
$\substack{n \geqslant m}$
$0, 0, ...)$. As before, we define A_k as $\underset{x}{E} [g(x) > y_k]$. But now $A_k \cap I$ may not be
right-complete. For example, it may happen that $f(1, 0, 0, 0, ...) > y_k$ but for every
other vector with 1 as first coordinate, f is $< y_k$.

Fortunately, we have $M_L(A_k \cap I) = M_U(A_k \cap I) = A_k$. First we prove that
$A_k \leqslant M_L(A_k \cap I)$. Let $(x_1, x_2, ...) \epsilon A_k$. Then $f(x_1, ..., x_n, 0, 0, ...) > y_k$ for at least
one $n \geqslant m$. Then for every positive integer r, $(x_1, ..., x_n, ..., x_{n+r}, 0, 0, ...) \epsilon A_k$
and so $\epsilon A_k \cap I$. Hence $(x_1, x_2, ...) \epsilon M_L(A_k \cap I)$. So $A_k \subseteq M_L(A_k \cap I)$. Now suppose
$(x_1, x_2, ...) \epsilon M_U(A_k \cap I)$. So there is an $n > m$ such that $(x_1, ..., x_n, 0, 0, ...) \epsilon A_k \cap I$.
So at least one of the numbers $f(x_1, ..., x_m, 0, 0, ...), f(x_1, ... x_{m+1}, 0, 0,), ..., f(x_1, ..., x_n,$
$0, 0, 0, ...)$ must be $> y_k$. Hence $(x_1, x_2, ...) \epsilon A_k$. Thus $M_U(A_k \cap I) \subseteq A_k$.

Since $M(A_k \cap I) = A_k$ and conditions G and H hold, we have by the Magni-
fication Theorem given in Section 3, $\delta(A_k \cap I) = P(A_k)$. The rest of the proof is as
before.

6. Application to Logarithmic Density

We now take up an important concrete case where Postulates (A) to (F) and
conditions G and H hold. We recall that if A is any set of positive integers,

$$\lim_{k \to \infty} \sup \frac{\left\{\underset{\substack{n \leqslant k \\ n \epsilon A}}{\Sigma} \frac{1}{n}\right\}}{\log k} \quad \text{and} \quad \lim_{k \to \infty} \inf \frac{\left\{\underset{\substack{n \leqslant k \\ n \epsilon A}}{\Sigma} \frac{1}{n}\right\}}{\log k}$$

are respectively called the upper and lower logarithmic densities of A. If they are equal,
this value is called the logarithmic density of A. If $L_L(A)$, $L_U(A)$, $N_L(A)$ and $N_U(A)$
are respectively the lower logarithmic, upper logarithmic, lower natural and upper
natural densities of A, $N_L(A) \leqslant L_L(A) \leqslant L_U(A) \leqslant N_U(A)$.

Let $q_1 = 2, q_2, q_3, ...$ be the prime numbers in ascending order of magnitude.
Consider the space X_n, $n = 1, 2, 3, ...$. Let $p_n^r (r = 0, 1, 2, ...)$ carry measure
$\frac{1}{q_n^r} \left(1 - \frac{1}{q_n}\right)$. In the space X we introduce the product measure. Let $(x_1, ..., x_{n_r}$
$0, 0, ...) \epsilon I$; with this vector we associate the positive integer $2^{x_1} 3^{x_2} ... q_n^{x_n}$. If $S \subseteq I$,
we define $\delta^U(S)$ to be the upper logarithmic density of the set of positive integers cor-
responding to the vectors in S. It is easy to verify that this δ^U satisfies Postulates

(A) to (F) and that the measure that arises in X coincides with the product measure we have formed. Condition G is true in this case. (This is true even for natural density).

We now verify condition H. Let B be a right-complete subset of I. We shall prove that B has logarithmic density equal to $P[M(B)]$. Since logarithmic density is finitely additive, the result is immediate if B has only a finite number of basic vectors. So we suppose that B has infinitely many basic vectors. We shall employ the Theorem: (Wintner, 1944, β^*, p, 53): A set S of positive integers has logarithmic density, $L(S)$, if and only if $(s-1)\sum\limits_{n\,\epsilon\,S}\dfrac{1}{n^s}$ tends to a limit as $s\to1+0$; in which case the limit is $L(S)$.

Let $\xi_1, \xi_2, \ldots$ be the basic vectors in B. Let $\xi_r = (x_1, x_2, \ldots, x_n, 0, 0, 0, \ldots)$, $x_n \geqslant 1$. $\sum\dfrac{1}{n^s}$, where each n corresponds to ξ_r, is

$$\zeta(s)\,\frac{\left(1-\dfrac{1}{2^s}\right)\cdots\left(1-\dfrac{1}{q_n^s}\right)}{\left(2^{x_i}\cdots q_n^{x_n}\right)^s},$$

this being true for all $s > 1$.

Thus
$$(s-1)\sum\limits_{n\,\epsilon\,B}\frac{1}{n^s}=(s-1)\zeta(s)\Big[\sum\limits_r f_r(s)\Big]$$

where
$$f_r(s)=\frac{\left(1-\dfrac{1}{2^s}\right)\cdots\left(1-\dfrac{1}{q_n^s}\right)}{\left(2^{x_1}\cdots q_n^{x_n}\right)^s}.$$

Since $\lim\limits_{s\to1+0}(s-1)\,\zeta(s)=1$, we now have to prove that

$$\lim\limits_{s\to1+0}\sum\limits_r f_r(s)=\sum\limits_r f_r(1)$$

this latter sum being the measure of $M(B)$.

Since
$$\sum\limits_{n\,\epsilon\,B}\frac{1}{n^s}=\zeta(s)\Big[\sum\limits_r f_r(s)\Big],\ \ \sum\limits_r f_r(s)\leqslant 1.$$

$\sum\limits_{r=1}^{\infty} f_r(s)$ is uniformly convergent on $(1+\delta)\leqslant s\leqslant 2$ for every $\delta > 0$; this is seen by noting that

$$f_r(s) < \frac{1}{(2^{x_1}\cdots q_n^{x_n})^{(1+\delta)}}\ \ \text{if } 1+\delta\leqslant s\leqslant 2.$$

Let us put $g_N(s)=\sum\limits_{r=1}^{N} f_r(s)$, $g(s)=\sum\limits_{r=1}^{\infty} f_r(s)$. Then $0\leqslant g_N(s)\leqslant 1$; $\lim\limits_{s\to1+0} g(s)\geqslant\lim\limits_{s\to1+0} g_N(s)=g_N(1)$ and hence $\varvarlim\limits_{s\to1+0} g(s)\geqslant g(1)$. Also, $g_N(1) < g(1)$.

We are now ready to prove that $g(s) \to g(1)$ as $s \to 1+0$. Suppose $\overline{\lim_{s \to 1+0}} g(s)$ is $> g(1)$. We recall that $g_N(s)$ converges to $g(s)$ uniformly on $[1+\delta, 2]$ for every $\delta > 0$. Let us choose a value s_1 of s very close to 1 such that $g(s_1)$ is almost equal to $\overline{\lim_{s \to 1+0}} g(s)$. Let us then choose a large N so that $g_N(s_1)$ is almost equal to $g(s_1)$. Now $g_N(1)$ is $< g(1) < \overline{\lim_{s \to 1+0}} g(s)$ and $g_N(s_1)$ is almost equal to this latter value. Thus we can find an s_2 exceeding 1 by an arbitrarily small quantity and an N such that $g'_N(s_2)$ is arbitrarily large (and positive). Let us now make this reasoning precise. Let ϵ be any fixed positive number $< \frac{1}{2} \{ \overline{\lim_{s \to 1+0}} g(s) - g(1) \}$; let K be any arbitrarily large positive number, let s_1 be any number > 1 such that $g(s_1) > \overline{\lim_{s \to 1+0}} g(s) - \epsilon$. Let $N(s_1) = N$ be so large that $|g(s_1) - g_N(s_1)| < \epsilon$. Then $g_N(s_1) > \overline{\lim_{s \to 1+0}} g(s) - 2\epsilon$. So by the mean-value theorem, there is an s_2 such that $1 < s_2 < s_1$ and

$$g'_N(s_2) = \frac{g_N(s_1) - g_N(1)}{s_1 - 1} > \frac{\overline{\lim_{s \to 1+0}} g(s) - g(1) - 2\epsilon}{(s_1 - 1)}.$$

By moving s_1 sufficiently close to 1 we can make this ratio $> K$. So $g'_N(s_2) > K$.

We shall now see that there is an absolute constant A such that $g'_N(s) \leqslant A$ for all N and all $s\epsilon(1, 2)$. In fact,

$$f'_r(s) = f_r(s) \left\{ \sum_{m=1}^{n} \frac{\log q_m}{(q_m^s - 1)} - \log (2^{x_1} \ldots q_n^{x_n}) \right\}$$

$$\leqslant f_r(s) \left\{ \sum_{m=1}^{n} \frac{\log q_m}{q_m - 1} - \log q_n \right\}, \text{ since } x_n \geqslant 1.$$

Now an elementary Theorem (Hardy and Wright, 1954, p. 348) in the analytic theory of numbers states that there is an absolute constant C such that for all n

$$\log q_n - C < \sum_{m=1}^{n} \frac{\log q_m}{(q_m - 1)} < \log q_n + C.$$

So

$$f'_r(s) \leqslant f_r(s) C.$$

$$\therefore \quad g'_N(s) = \sum_{r=1}^{N} f'_r(s) \leqslant C \, g_N(s) < C.$$

This contradiction proves that $\lim_{s \to 1+0} g(s) = g(1)$.

It may be noted that condition H no longer holds if, in the preceding discussion, logarithmic density is replaced by natural density. In fact, Besicovitch (1934, p. 336-341) has constructed a set of positive integers such that the set S of their multiples has no natural density; but S is a right-complete set.

This paper is based on a dissertation submitted in partial fulfilment of the requirements for the Ph.D. degree at the University of Illinois, in 1960. The author wishes to express his gratitude to Professor J. L. Doob for his guidance during the preparation of the thesis.

REFERENCES

Besicovitch, A. S. (1934) : On the density of certain sequences of integers. *Math. Ann.*, **110**, 336-341.

Hardy, G. H. and Wright, E. M. (1954) : *An Introduction to the Theory of Numbers*, (Third Edition), Oxford.

Kolmogorov, A. N. (1956) : *Foundations of the Theory of Probability*, Chelsea Publishing Company, New York.

Wintner, Aurel (1944) : *The Theory of Measure in Arithmetical Semigroups*, Baltimore.

ON STRUCTURE, RELATION, Σ, AND EXPECTATION OF MEAN SQUARES*

By GEORGE ZYSKIND

Iowa State University

SUMMARY. Some properties of balanced population structures are investigated. The forms for the expected value of squares of balanced sample means from balanced population structures are obtained, and this is done also for the type of sample mean arising in a class of randomized experiments. The canonical form of the Σ expansion of the expected value of the square of the sample mean is demonstrated. The form of the expected value of mean squares in the analysis of variance for a large class of situations is then derived as a consequence.

1. INTRODUCTION

The present paper concerns itself with a rather general approach to certain basic aspects of questions of experimental design connected with the technique of the analysis of variance.

The mathematical representation of common experimental designs is generally considered to be covered by various special cases of the general linear hypothesis theory in its current formulation. Though that theory has been extremely successful and is very useful, some of its drawbacks have of late been considered sufficiently important to warrant an approach not covered by it at present. Thus, Kempthorne (1952) explicitly introduces randomization variables in order that the mathematical representation related to the designs he considers reflect a one-to-one correspondence with the way the experiments are to be carried out. This approach was insisted upon in later publications by Kempthorne (1955), Wilk (1955a and 1955b) and Wilk and Kempthorne (1955, 1956a, 1956b, 1957). Further, to strengthen the foundations of their approach and also to explain their position in connection with the "mixed model controversy" these authors decided to use "derived linear" rather than "assumed linear" models in all the particular problems considered. One characteristic of "derived" models is that their application does not require any assumption concerning the form of the response as a function of the values of the factors influencing it.

In the formulation of a derived statistical model for a particular experimental situation, one first specifies the "population identity." The population identity expresses the typical "actual" or "conceptual" response as a sum of 'population components' each of which is a relevant linear function of the possible actual or conceptual numbers yielded by the experiment. The components are constructed so as to bear a high degree of correspondence with the usual main effect and interaction terms of assumed linear models. They are precisely defined, but all that is required for their construction, as indeed for the construction of the identity when balance obtains, is the specification of the "population structure" with regard to the relationship of "nesting" among the set of individual entities envisaged to possibly influence the observed response.

In the exploration of results obtained in his thesis Wilk (1955a) found that introduction of certain well-defined linear functions of the usual components of variation, σ^2's, substantially simplified the form of expectations of mean squares in all the analysis of variance tables he considered. Wilk denoted these quantities by Σ's (read as cap sigmas) suggesting that under symmetric conditions extensions of the results on expectations of mean squares were implicit in the forms obtained for the cases he studied. The paper on non-additivities in a Latin Square design by Wilk and Kempthorne (1957), exemplifies the types of Σ expressions involved.

In the present paper we discuss the notion of relation and structure in experimental design and give, in the notation here introduced, the general definition of the Σ quantities. The main result of the paper concerns the simple general Σ form of expected values of squares of typical observational means involved in analysing experiments. The simple Σ form of expected mean squares in the analysis of variance follows then as a direct consequence of the fact that in "balanced" cases each such expectation can be written out uniquely and in an easily specified manner as a linear function of expected values of squares of the typical observational means. As the contents of the present paper makes clear, the applicability of the above statements to physical problems is, indeed, very broad.

2. POPULATION STRUCTURES

Consider a response designated by Y. Suppose it is envisaged to depend entirely on a finite number of entities, e.g., pressure, temperature, etc., every one of which is indicated by a corresponding subscript in the notation for the response, where the range of these subscripts is over the possible levels of the entity in question. We restrict ourselves to situations in which every combination of the levels of subscripts is admissible. The physical layout and character of the entities in question are usually such as to admit a natural set-up with respect to the relation of hierarchal arrangements of the entities. An entity is said to be hierarchal, or nested, within another set of entities, S say, if the unique identification of any one of its elements requires also the specification of some particular set of elements of the entities S and the entities S are said to nest it. For example, in the structure of the experimental material of the randomized block design, the unique identification of a particular plot requires not only the plot number but also the specification of the block containing the plot. Thus, if we denote a typical possible response in the randomized block by Y_{ijk}; where the index i refers to the block classification, the index j to plot, and the index k to the treatment classification then the unique specification of a plot requires always not only the particular value of the index j but also the special value of the index i appropriate to the block in which the plot is nested. Symbolically, the structure of the entities involved in a typical conceptual response of a randomized block may be expressed as $(i : j)(k)$, where the brackets separate the different types of entities and the colon indicates that the unit entity, j, is nested in the block entity i. According to our notation it would make no sense to write $j : i$ in this example.

A more complex structure of the fundamental entities is exemplified by the case where the experimental material is stratified into sources, S, with each source being cross-classified by rows, R, and columns, C. Suppose further that there are two types of treatment factors, say A and B, but that repeated attempts at the same level of a treatment are subject to errors in realization. A conceptual response is, therefore, due to both the levels aimed at and a deviation from it, or a sublevel. The treatment structure then is hierarchal—sublevels within levels. If we denote sublevels of A and B by a and b, respectively, then a symbolic representation of the structure of the entities involved is

$$(S:RC) \quad (A:a) \quad (B:b).$$

An extension of the above example is furnished by the case where the columns within sources are further subdivided into strips, denoted by L. The rows do not nest the strips and hence the complete representation of the structure of the entities involved requires sub-brackets within the main bracket for the experimental material. The symbolic representation of the situation is

$$(S:(R)(C:L)) \quad (A:a) \quad (B:b).$$

Finally, consider the following illustrative example. Four entities are involved in the structure. Denote them by P, Q, S, and R. The set of relations is: Q nested in S, and R nested in SP combinations.

Symbolically, $\qquad (S:Q)(P) \quad$ and $\quad (SP:R).$

One can obtain partial population means by averaging over the entire range of values of particular sets of subscripts. Partial means are denoted by the usual symbol for a response but with omission of subscripts over which the average has been taken. An admissible mean is defined as one in which whenever a nested index appears then all the indices which nest it appear also. Our considerations are restricted to admissible means only. The indices of an admissible partial mean which nest no other indices of that mean are said to constitute the set of indices belonging to the rightmost bracket. It is convenient to indicate the grouping of the indices of the rightmost bracket by using parenthesis, ( ), and also to group in this way other sets of indices when we wish to emphasize that for some structural reason they belong to the same category. Thus, in our randomized block case the admissible partial means are six in number and may be denoted by: Y, Y_i, Y_k, $Y_{(ik)}$, $Y_{(i)(j)}$, $Y_{(i)(jk)}$.

From every partial mean linear combinations of means can be formed which are of special physical and formal significance. These linear combinations, henceforth called components, are obtained by selecting all those partial means which are yielded by the mean in question when some, all, or none of its rightmost bracket subscripts are omitted in all possible ways. Whenever an odd number of indices is omitted the mean is to be preceded by a negative sign, whenever an even number is omitted the mean is to be preceded by a positive sign. The number zero is considered even. For example, in the randomized block the partial mean $Y_{(i)(j)}$ leads to the component $(Y_{(i)(j)} - Y_i)$, the mean $Y_{(i)(jk)}$ to $(Y_{(i)(jk)} - Y_{(i)(j)} - Y_{ik} + Y_i)$, and the

mean Y to (Y). The components thus constructed have a correspondence with the effects and interactions of the usual assumed linear models. They also bear a correspondence to the various terms of a Taylor expansion of well-behaved functions.

The following facts are immediate consequences of the definitions.

(1) If the rightmost group of the leading term of a component consists of p indices then the number of means present in the component is 2^p. This follows from the fact that the number of means in the component is equal to the total number of groups of size zero to p inclusive that can be formed from p different objects. Since each object can be dealt with in exactly two ways, it can be included in a group or excluded from it, the total number of groups is 2^p.

(2) The sum of the coefficients of means in any component is zero excluding the coefficients of the component Y for which the sum is one. This is so because coefficients plus one or minus one are assigned to partial means according to an "even-odd" criterion. Adding the signed coefficients of the partial means we get

$$\binom{p}{p} - \binom{p}{p-1} + \binom{p}{p-2} - \binom{p}{p-3} + \ldots + (-1)^p \binom{p}{0}$$

$$= (1+x)^p \qquad \text{for } x = -1$$

$$= (1-1)^p = 0 \text{ for } p \neq 0 \qquad \text{q.e.d.}$$

We now show that for any given population structure the typical response can be expressed identically as a sum of all its corresponding components. This relation is called the population identity. To establish the identity consider any one admissible partial mean. The partial mean appears in the expanded form of those components for which the subscripts of the leading terms exceed or are identical with the subscripts of the partial mean, and for which the excess subscripts appear only in the rightmost bracket of the leading term. It follows that all the qualifying leading means can be obtained from the partial mean by adding to its indices none, some, or all of the indices of a particular finite set. Further, since whenever the number of excess indices is even the corresponding component contains the partial mean considered with coefficient plus one; and whenever it is odd with coefficient minus one we see from the argument developed above that for any partial mean, except the one containing all indices of an individual response, the sum of its coefficients over all the components is zero. For the exceptional term the coefficient is one. This completes the proof that the decomposition of an individual conceptual response into a sum of components is identical. q.e.d.

As an illustration we notice that the term Y_i in the randomized block case appears in, and only in, components whose leading terms have none of, one of, or both of j and k as an excess over i in their rightmost group. The number of such terms is $2^2 = 4$. They are $Y_i, Y_{ik}, Y_{i(j)}, Y_{i(jk)}$, and their corresponding components are

$$(Y_i - Y), \ (Y_{ik} - Y_i - Y_k + Y),$$
$$(Y_{(i)(j)} - Y_i), \ (Y_{(i)(jk)} - Y_{(i)(j)} - Y_{ik} + Y_i).$$

We further note that the sum of the Y_i's over these components is zero, as it should be.

The population components can be shown to satisfy some very simple and useful relations when the range of any one subscript of the population structure is the same for every particular set of values of the other subscripts. When this condition obtains the population structure is said to be balanced. Because of reasons of mathematical simplicity and because structures involved in experimental designs are generally chosen to be balanced, the population structures discussed in the present paper will be taken to satisfy the balance condition.

We now state a number of basic assertions about balanced population structures. Details of proofs which are omitted can be found in Zyskind (1958).

(1) For any type of component the sum of values of the components is zero over the population range of any one index of the rightmost bracket of the leading term of the type.

Thus, the above statement generalizes identities such as $\sum\limits_{i} (Y_i - Y) = 0$.

(2) The sum of squares of the responses over all values of all indices is equal to the sum of squares of all the individual typical components over the ranges of these same indices.

The above assertion generalizes identities such as

$$\sum\limits_{ij} Y^2_{(i)(j)} = \sum\limits_{ij} Y^2 + \sum\limits_{ij} (Y_i - Y^2) + \sum\limits_{ij} (Y_{(i)\,(j)} - Y_i)^2.$$

Proof: Denote the value of a component by a capital letter and subscripts, using identical subscripts with those of the leading term of the component. The product of the values of two individual components of different types involves an index which is in the rightmost bracket of one of the components but not at all among the indices of the other component. By assertion one the sum of products of these two types of components over such an index is zero. Hence, the sum of products of components of different types over all the indices of the population is zero. The validity of the assertion follows as a consequence. q.e.d.

(3) The total number of linearly independent components is exactly N where N is the total number of possibly different responses.

(4) The number of linearly independent values of a given type of component is equal to the product of the population ranges of indices of the component, which do not belong to its rightmost bracket times the product of the diminished ranges of the indices of the rightmost bracket. (By the range here is meant the number of distinct values taken on by the index in question; the diminished range equals the range minus one). Further, the sum of these numbers over the components of different types is equal to N.

(5) For every type of component the sum of squares of values of the component over the ranges of all the indices of the rightmost bracket is equal to the sum over the same set of indices of a linear function of squares of partial means making

119

up the component, with the coefficients of the squares being the same as those defining the component in terms of corresponding partial means.

The above assertion generalizes identities such as

$$\sum_{jk} (Y_{(i)(jk)} - Y_{ij} - Y_{ik} + Y_i)^2 = \sum_{jk} (Y_{(i)(jk)}^2 - Y_{ij}^2 - Y_{ik}^2 + Y_i^2),$$

for all values of i.

Proof: For a particular type of component consider the sum of squares of its values over all the indices of the rightmost bracket. If there are p indices of the rightmost bracket, then there are 2^p subsets of these; these subsets are in one-to-one correspondence with the partial means appearing in the initial and final expressions. To establish the identity one first expands the square and puts the resulting square and cross-product terms into 2^p categories, the category being specified by the set of rightmost bracket indices which the two factors of the product have in common. Such sets will be called 'intersection sets.' Further, for each product we define the 'excess set' to consist of those indices which appear in exactly one of the two factors. The coefficient of any product is then $(-1)^i$ where i is the number of indices in its excess set. Because of balance, when any product is summed over its excess set one obtains the square of the partial mean corresponding to the intersection set times the product of the ranges of the excess indices. The product of ranges can then be replaced by summation over the excess indices thus restoring the summation over all rightmost bracket indices. It remains to combine like terms. Let p, q, i denote respectively the number of indices in the rightmost bracket, the intersection set, the excess set. For fixed p, q one has $i = 0, 1, ..., p-q$. All possible combinations of excess indices will appear in the expansion for each i, these are $\binom{p-q}{i}$ in number. Further, for any particular combination the i indices may be divided between the factors in 2^i ways. Thus for q indices in the intersection set, i.e., q indices in the partial mean, the coefficient of the squared partial mean is

$$. \sum_{i=0}^{p-q} \binom{p-q}{i}(-2)^i = \begin{cases} (1+x)^{p-q}, \text{ for } x = -2 \\ (1-2)^{p-q} = (-1)^{p-q} \\ \quad 1 \text{ if } p-q \text{ is even,} \\ -1 \text{ if } p-q \text{ is odd.} \end{cases}$$

and

Thus, the coefficient of the square of the partial mean is the same as the coefficient of that mean in the definition of the component. q.e.d.

Corollary : *The type of identity specified by assertion 5 is valid when summation takes place over the ranges of all the indices of a typical response.*

Exploitation of identities specified by the corollary forms the basis of our approach to finding expected values of mean squares in the analysis of variance table.

Definition : The number of linearly independent values of components of a type is said to be the number of degrees of freedom of the type of component. Also, the number of linearly independent possible responses is said to be the number

of degrees of freedom of the set of possible responses. We shall often abbreviate the expression degrees of freedom by writting D.F.

For a given type of component the number of degrees of freedom is as stated in assertion 4.

Definition : The sum of squares of values of components of a given type over all the population ranges of the indices used to denote the component divided by the number of degrees of freedom of the type is said to be the component of variation of the type of component.

Henceforth we denote particular components of variation by σ^2's with subscripts, bracketed into groups, corresponding to the subscripts of the types of components to which the σ's refer.

We now define the Σ's.

Definition: Consider a particular type of component and all σ^2's of the following form

(i) the set of subscripts of σ^2 includes the set of subscripts corresponding to the leading term of the component as a subset.

(ii) the excess subscripts belong exclusively to the rightmost bracket of σ^2.

The linear combination of all such σ^2's, where the coefficient of a particular σ^2 with k excess subscripts is

$$(-1)^k \frac{1}{\text{product of population ranges of the excess indices}}$$

is defined as the Σ corresponding to the type of component under consideration. The subscript notation for the Σ is to be the same as for the type of component.

It should be pointed out that the component of variation corresponding to the null set is $Y^2 = \sigma^2_{(\theta)}$, and that the corresponding $\Sigma_{(\theta)}$ is uniquely defined. The introduction of $\Sigma_{(\theta)}$ is of prime importance to the development of the present approach.

To fix ideas consider randomized blocks with B blocks each of P plots, and with T treatments. The Σ's are as follows

$$\Sigma_{(\theta)} = \sigma^2_{(\theta)} - \frac{1}{B}\sigma^2_{(B)} - \frac{1}{T}\sigma^2_{(T)} + \frac{1}{BT}\sigma^2_{(BT)}$$

$$\Sigma_{(B)} = \sigma^2_{(B)} - \frac{1}{P}\sigma^2_{(B)(P)} - \frac{1}{T}\sigma^2_{(BT)} + \frac{1}{PT}\sigma^2_{(B)(PT)}$$

$$\Sigma_{(T)} = \sigma^2_{(T)} - \frac{1}{B}\sigma^2_{(BT)}$$

$$\Sigma_{(BT)} = \sigma^2_{(BT)} - \frac{1}{P}\sigma^2_{(B)(PT)}$$

$$\Sigma_{(B)(P)} = \sigma^2_{(B)(P)} - \frac{1}{T}\sigma^2_{(B)(PT)}$$

$$\Sigma_{(B)(PT)} = \sigma^2_{(B)(PT)}.$$

121

As another example we may verify that for the structure specified by

$$(S : Q)(P) \text{ and } (SP : R),$$

where the respective population ranges are S, P, Q and R, the set of Σ's is

$$\Sigma_{(\theta)} = Y^2 - \frac{1}{S}\,\sigma^2_{(S)} - \frac{1}{P}\,\sigma^2_{(P)} + \frac{1}{SP}\,\sigma^2_{(SP)}$$

$$\Sigma_{(S)} = \sigma^2_{(S)} - \frac{1}{P}\sigma^2_{(SP)} - \frac{1}{Q}\,\sigma^2_{(S)(Q)} + \frac{1}{PQ}\sigma^2_{(S)(PQ)}$$

$$\Sigma_{(P)} = \sigma^2_{(P)} - \frac{1}{S}\,\sigma^2_{(SP)}$$

$$\Sigma_{(SP)} = \sigma^2_{SP} - \frac{1}{Q}\,\sigma^2_{(S)(PQ)} - \frac{1}{R}\sigma^2_{(SP)(R)} + \frac{1}{QR}\,\sigma^2_{(SP)(QR)}$$

$$\Sigma_{(S)(Q)} = \sigma^2_{(S)(Q)} - \frac{1}{P}\,\sigma^2_{(S)(QP)}$$

$$\Sigma_{(S)(PQ)} = \sigma^2_{(S)(PQ)} - \frac{1}{R}\,\sigma^2_{(SP)(QR)}$$

$$\Sigma_{(SP)(R)} = \sigma^2_{(SP)(R)} - \frac{1}{Q}\sigma^2_{(SP)(QR)}$$

$$\Sigma_{(SP)(QR)} = \sigma^2_{(SP)(QR)}.$$

The analysis of variance, introduced by R. A. Fisher in 1918, is a technique which lays out in tabular form the breakdown of the total sum of squares into separate parts, each of which can usually be given a supposedly physical meaning in the sense that it describes an assignable source (or complex of sources) of variation.

Corresponding to any problem to which the application of the analysis of variance technique is appropriate there exist at least two major types of analyses of variance—one for the population and one for the observed sample. For the purposes of this paper we define the population analysis of variance to be the tabular partitioning of the total sum of squares and degrees of freedom of all the responses of the population into parts, each part corresponding to one type of component. The validity of assertions 2 and 4 ensures that construction of such a table is possible for every balanced population structure.

For each part we shall also exhibit the quotient of the sum of squares by the degrees of freedom, and we shall call the result the mean square of the part. Since the different quantities associated with a part are usually exhibited in an orderly fashion in a single line, we shall use the terms part and line interchangeably.

As an immediate consequence of our definitions we see that for any particular line:

Mean Square = (product of population ranges of indices not involved in the type
of component corresponding to the line)×

$$\underset{\substack{\text{indices of} \\ \text{component}}}{\Sigma}\ (\text{component})^2/\text{number of degrees of freedom}$$

=(number of individual responses entering into a typical leading mean
of the component)$\times\sigma^2_{\text{component}}$

3. Balanced sample means from balanced population structures

In the mathematical models for the experimental designs we consider, we take care that the models reflect a one-to-one correspondence with both the initial population structure and the physical way in which the experiment is to be performed. The use of design and sampling random variables, which we shall illustrate for specific instances later, has proved extremely valuable to this end. Since in our formulation we use derived or definitional population models our method in its initial general formulation does not depend on any special assumptions. Thus, it has the advantage of allowing us to introduce simplifying assumptions only as they are needed, and so to explore in detail just how far a minimal set of assumptions will carry us. Finally, because in the present approach we always initially consider finite population structures, fixed, random, and mixed situations come out as particular and usually simple cases of the general formulation.

The details of the mathematical procedure we employ are as follows. The conceptual counterpart of the samplings and random assignments of chosen entities involved in the carrying out of the various experimental schemes can be obtained by conceiving of carrying out the similar operations in the population of index values of all the possible responses. This is accomplished by the explicit use of the sampling and design random variables. The statistical model for a sample observation is then one indicating explicitly the physical process by which the given observation was or is to be selected from the set of population values. From the statistical model for an arbitrary experimental observation it is easy to see that every partial sample mean is expressible as a sum of sample means of possibly different types from the various population components. Thus, special groupings of sample observations induce particular types of samples from the population components.

An example illustrating the situation is that of the simple one-fold hierarchal or nested population. Here the typical population observation, Y_{ij}, is expressed identically in terms of components as follows

$$Y_{ij} = Y + (Y_i - Y) + (Y_{(i)(j)} - Y_i).$$

Any sample observation or any mean of sample observations will involve clearly a sample of the $(Y_i - Y)$'s and a sample of the $(Y_{(i)(j)} - Y_i)$'s. The particular samples of the $(Y_i - Y)$'s and $(Y_{(i)(j)} - Y_i)$'s involved depend on the sample of the Y_{ij}'s actually chosen and are said to be induced by it.

We shall denote a particular sample observation by the symbol x where the subscripts of x indicate the sampling orders, in terms of the various population classifications, in which the sample value was obtained. For reasons of mathematical difficulty and also primary interest, we restrict our attention in the remainder of the paper to types of samples which will henceforth be called balanced.

Definition: A sample is said to be balanced with respect to all subscripts used in the representation of an arbitrary sample observation if the sample range of

any one of the subscripts is the same for every set of particular values the other subscripts may assume.

In what follows we exhibit in order of increasing complexity of structure and derivation results on the expectation of squares of sample means. We also define, corresponding to any balanced sampling scheme, the sample analysis of variance. Making use of the fact that the mean square of any line of the table can be written as a known linear function of squares of sample means we state and prove, in both the σ^2 and Σ languages, the general theorem about the simple form of the expected value of the mean square of any line.

Consider a random sample of size n from the population of N elements whose values are denoted by Y_i, $i = 1, 2, ..., N$. Denote the population mean of the Y_i's by Y. Denote also by $x_{i\cdot}$, $i^\cdot = 1, 2, ..., n$, the value of $x_{i\cdot}$-th observed sample member in order of selection. Define now nN random variables as follows

$\alpha_i^{i^\cdot} = 1$ if $i^\cdot$-th chosen item in the sample is the i-th item in the population.

$\alpha_i^{i^\cdot} = 0$ otherwise.

Each of the selection variables has the following simple properties

$$P(\alpha_i^{i^\cdot} = 1) = \frac{1}{N}; \quad P(\alpha_i^{i^\cdot} = 0) = 1 - \frac{1}{N};$$

$$E(\alpha_i^{i^\cdot}) = E(\alpha_i^{i^\cdot})^2 = \frac{1}{N}, \qquad \text{for all } i^\cdot, i.$$

$$E(\alpha_i^{i^\cdot} \alpha_{i'}^{i^{\cdot\prime}}) = \frac{1}{N} \times \frac{1}{N-1}, \quad \text{for } i \neq i', \ i^\cdot \neq i^{\cdot\prime}.$$

Here, and in the remainder of the paper the symbol P is used as a shorthand for the word probability and the symbol E denotes the expectation operator.

In terms of the elementary random variables the value of any $x_{i\cdot}$ may be written

$$x_{i\cdot} = \sum_{i=1}^{N} \alpha_i^{i^\cdot} Y_i = \sum_{i=1}^{N} \alpha_i^{i^\cdot} \left[Y + (Y_i - Y) \right]$$

$$= \mu + \sum_{i=1}^{N} \alpha_i^{i^\cdot} A_i,$$

where $\mu = Y = \dfrac{\sum_{i=1}^{N} Y_i}{N} = $ population mean, $A_i = Y_i - Y$.

Hence $\quad E\, x_{i\cdot} = \mu + \dfrac{1}{N} \sum_i (Y_i - Y) = \mu$, since $\sum_{i=1}^{N} (Y_i - Y) = 0$.

Further, $E x_{i*}^2 = E(Y + \sum_i \alpha_i^{i*}(Y_i - Y))^2 = Y^2 + E \sum_i \alpha_i^{i*}(Y_i - Y)^2 = Y^2 + \frac{1}{N} \sum_{i=1}^{N} (Y_i - Y)^2$

$$= Y^2 + \left(1 - \frac{1}{N}\right) \sigma_{(A)}^2 \left(\text{ where } \sigma_{(A)}^2 = \sum_i \frac{(Y_i - Y)^2}{N-1}\right)$$

$$= \left(Y^2 - \frac{1}{N} \sigma_A^2\right) + \sigma_{(A)}^2 = \Sigma_{(\theta)} + \Sigma_{(A)}.$$

It follows that $\qquad V(x_{i*}) = E(x_{i*}^2) - (E(x_{i*}))^2 = \left(1 - \frac{1}{N}\right) \sigma_{(A)}^2 .$

Further, $\qquad\qquad x = \frac{1}{n} \sum_{i*=1}^{n} x_{i*} = \mu + \frac{1}{n} \sum_{i*=1}^{n} \sum_{i=1}^{N} \alpha_i^{i**} A_i .$

Hence, employing the properties of the sampling random variables we can show easily that

$$Ex^2 = \mu^2 + \left(1 - \frac{n}{N}\right) \frac{\sigma_A^2}{n} .$$

. We notice that the variance of the sample mean $= Ex^2 - \mu^2 = \left(1 - \frac{n}{N}\right) \frac{\sigma_{(A)}^2}{n} .$
This, of course, is a well-known finite sampling elementary result. We focus attention, however, on the facts that the square of a sample mean of size n from the population of the A_i's has expectation

$\left(1 - \frac{n}{N}\right) \frac{\sigma_{(A)}^2}{n}$, and that $Ex^2 = \mu^2 + \left(1 - \frac{n}{N}\right) \frac{\sigma_{(A)}^2}{n} = \left(\mu^2 - \frac{1}{N} \sigma_A^2\right) + \frac{\sigma_{(A)}^2}{n} = \Sigma_\theta + \frac{\Sigma_{(A)}}{n}.$

We have exemplified the use of sampling variables in the above simple illustration because these and also design random variables are used extensively in the paper in the formulation of derived linear models and also in the arguments of several of the underlying proofs.

The result $Ex^2 = \mu^2 + \left(1 - \frac{n}{A}\right) \frac{\sigma_{(A)}^2}{n} = \Sigma_{(\theta)} + \frac{\Sigma_{(A)}}{n}$ admits convenient analogues in more dimensions. Thus, consider elements with values P_{ij} arranged in a two-way table of A rows and B columns; $i = 1, 2, ..., A$ and $j = 1, 2, ..., B$; with the values P_{ij} subject to the condition:

$$\sum_{i=1}^{A} P_{ij} = \sum_{j=1}^{B} P_{ij} = 0.$$

A sample of elements is now chosen as follows. A random selection of a out of A rows is made, and independently a random selection of b columns out of B; the elements of the sample are the elements of the intersections of the chosen rows and columns. A sample drawn according to the above procedure is said to be a cross sample. The term "bisampling" was used by Cornfield and Tukey (1956) but we do not favour it

because its use would necessitate terms like "trisampling" and so on. A schematic representation of such a sample is given in Figure 1 below.

entity B
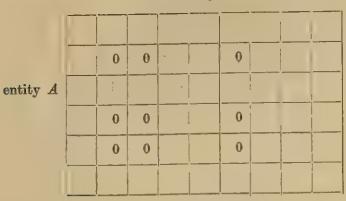
Fig. 1 : A cross sample from a cross two-dimensional population.

The physical situation can again be described by introducing sampling random variables. Let

$\alpha_i^{i^*} = 1$ if the i^*-th choice from among the i's selects all P_{ij}'s having the subscript i,

and $\quad \alpha_i^{i^*} = 0$ otherwise.

$\beta_j^{j^*} = 1$ if the j^*-th choice from among the j's selects all P_{ij}'s having the subscript j.

and $\quad \beta_j^{j^*} = 0$ otherwise,

An observed sample value, denoted by x_{i*j*}, is then expressed as a function of the P_{ij}'s as

$$x_{i*j*} = \sum_{ij} \alpha_i^{i^*} \beta_j^{j^*} P_{ij}.$$

The expected value of the square of the sample mean, x, can then easily be shown to be

$$Ex^2 = E\left(\frac{1}{ab} \left(\sum_{i*j*} x_{i*j*} \right) \right)^2 = \left(1 - \frac{a}{A} \right)\left(1 - \frac{b}{B} \right) \frac{\sigma_{AB}^2}{ab},$$

where $\quad \sigma_{AB}^2 = \dfrac{\sum\limits_{ij} P_{ij}^2}{(A-1)(B-1)}.$

Thus, a multiplicative correction factor of the form shown above appears corresponding to every classification involved, and σ_{AB}^2 is to be divided by the total sample size.

The condition $\sum\limits_{i=1}^{A} P_{ij} = \sum\limits_{j=1}^{B} P_{ij} = 0$ is one which is satisfied by the component types of balanced population structures. It is *not* an assumed property.

It should be noticed that whenever complete selection of at least one of the classifications is made, then the general formula yields the value zero for Ex^2. This is in agreement with the fact that under the above condition x is equal to zero

identically. The condition is seen to correspond to the situation when the classification in question is termed to be fixed in common statistical terminology.

The connection with the usual two-way balanced classification is as follows. Let Y_{ij}, $i = 1, 2, ..., A$; $j = 1, 2, ..., B$, be the typical response of the ij-th cell. Then Y_{ij} admits an identical decomposition

$$Y_{ij} = Y + (Y_i - Y) + (Y_j - Y) + (Y_{ij} - Y_i - Y_j + Y) = \mu + A_i + B_j + AB_{ij}.$$

If the sample is drawn as described above then the i^*j^*-th observed sample value is related to the Y_{ij}'s and to the population components by

$$x_{i^*j^*} = \mu + \sum_i \alpha_i^{i^*} A_i + \sum_j \beta_j^{j^*} B_j + \sum_{ij} \alpha_i^{i^*} \beta_j^{j^*} (AB)_{ij},$$

$$Ex^2 = \mu^2 + \left(1 - \frac{a}{A}\right)\frac{\sigma_{(A)}^2}{\blacksquare} + \left(1 - \frac{b}{B}\right)\frac{\sigma_{(B)}^2}{b} + \left(1 - \frac{a}{A}\right)\left(1 - \frac{b}{B}\right)\frac{\sigma_{(AB)}^2}{ab}$$

$$= \Sigma_{(\theta)} + \frac{1}{a}\Sigma_{(A)} + \frac{1}{b}\Sigma_{(B)} + \frac{1}{ab}\Sigma_{(AB)}$$

where for example, $\qquad \sigma_{(A)}^2 = \dfrac{\sum\limits_{i=1}^{A} A_i^2}{A-1} = \dfrac{\sum\limits_{i=1}^{A}(Y_i - Y)^2}{A-1}$.

By introducing a little more notation it can be shown without much trouble, as is done by Zyskind (1958), that the basic form of the above results remains valid regardless of the number of crossed classifications involved. The fact that the Σ forms follow as a consequence of the σ^2 forms will be shown here later for situations of which the present one is a particular case.

Consider next a one-fold nested population of AB ordered items arranged into A groups of B items each. Denote the value of the j-th item of the i-th group by $Y_{ij}(= Y_{i(j)})$.

We have identically

$$Y_{ij} = Y + (Y_i - Y) + (Y_{i(j)} - Y_i) = \mu + A_i + B_{i(j)},$$

where for example $B_{i(j)} = Y_{ij} - Y_i$.

Since the population is a balanced one

$$\sum_{i=1}^{A} A_i = 0, \quad \sum_{j=1}^{B} B_{i(j)} = 0. \text{ for all values of } i.$$

Let the sampling procedure consist of choosing randomly a out of A groups and in each of the groups selected choosing randomly b out of the B elements.

Let $x_{i^*j^*}$ denote the value of the j^*-th chosen element in the i^*-th chosen group. We link the variable $x_{i^*j^*} (= x_{i^*(j^*)})$ to the values of the population items by defining elementary selection random variables of the type $\alpha_i^{i^*}$, $\beta_{i^*j}^{i^*j^*}$, where for example

$$\beta_{i^*j}^{i^*j^*} = \begin{cases} 1 \text{ if the } j^*\text{-th chosen unit of } i^*\text{-th chosen group is the} \\ \quad j\text{-th unit of that group} \\ 0 \text{ otherwise,} \end{cases}$$

$$j^* = 1, 2, ..., b; \quad j = 1, 2, 3, ..., B.$$

127

4

It follows now that

$$x_{i \bullet j \bullet} = \sum_{ij} \alpha_i^{i \bullet} \beta_{i \bullet j}^{i \bullet j \bullet} \ \ Y_{ij} = \sum_{ij} \alpha_i^{i \bullet} \beta_{i \bullet j}^{i \bullet j \bullet} \ (\mu + A_i + B_{i(j)})$$

$$= \mu + \sum_i \alpha_i^{i \bullet} A + \sum_{ij} \alpha_i^{i \bullet} \beta_{i \bullet j}^{i \bullet j \bullet} \ B_{i(j)}.$$

By making use of the properties of the random variables $\alpha_i^{i \bullet}$, $\beta_{i \bullet j}^{i \bullet}$, we can easily verify that

$$Ex_{i \bullet j \bullet}^2 = \mu^2 + \left(1 - \frac{1}{A} \right) \sigma^2_{(A)} + \left(1 - \frac{1}{B} \right) \sigma^2_{(A)(B)} = \Sigma_{(\theta)}^{\bullet} + \Sigma_{(A)} + \Sigma_{(A)(B)}.$$

Also,
$$Ex^2 = E \left(\frac{1}{ab} \sum_{i \bullet j \bullet} x_{i \bullet j \bullet} \right)^2 = \mu^2 + \left(1 - \frac{a}{A} \right) \frac{\sigma^2_{(A)}}{a} + \left(1 - \frac{b}{B} \right) \frac{\sigma^2_{(A)(B)}}{ab}$$

$$= \Sigma_{(\theta)} + \frac{1}{a} \Sigma_{(A)} + \frac{1}{ab} \Sigma_{(A)(B)}.$$

The generalizations of the formulas exhibited so far can without excessive trouble be shown to be valid. The detailed proofs are given by Zyskind (1958) and are here omitted for reasons of economy of space.

Definition : The finite correction factor of size j due to the index i is the number $\left(1 - \dfrac{j}{\text{population range of } i} \right)$ and is denoted by f_i^j.

We now summarize the bulk of our conclusions thus far.

Theorem 1 : *The expectation of the square of any admissible partial mean, arising from a balanced sample of observations, is equal to a linear function of all the different components of variation of the population. The coefficient of each component of variation is the product of the finite correction factors, one factor for each index of the rightmost bracket of the component and each of the form f_i^α where α is the number of different values of the index i entering into the partial mean, divided by the number of different values of the component entering into the formation of the partial mean in question.*

We notice that relative to a partial observational mean the sample size of index i is one whenever the index corresponding to i appears in the partial mean. This leads to the rule that the number of different components of a type entering into the formation of a mean is equal to the product of sample sizes of indices, which the type of component has in excess over those of the admissible partial mean. Also, $f_i^\alpha = 1 - \dfrac{1}{\text{range of index } i}$ if the indices corresponding to i appear in both the partial sample mean and in the rightmost bracket of the component of variation; if the indices corresponding to i appear in the rightmost bracket of the component of variation but not in the partial mean then $f_i^\alpha = 1 - \dfrac{\text{sample range of } i \ (= \alpha)}{\text{population range of } i}$.

We now seek to express the result of the above theorem in Σ form.

Consider the term involving a particular σ^2. Let X be the set of the non-rightmost subscripts of σ^2 and let Y be the set of rightmost subscripts. Let Z be an

arbitrary set of subscripts such that $Z \subseteq Y$. Denote the number of subscripts in Z by q and for $q = 0$ define the product of the ranges of the subscripts in Z to be one. Let $N_{X+(Y-Z)}$ denote the number of different components, whose type is specified by the set of indices $X+(Y-Z)$, entering into the formation of the partial sample mean. We carry out the argument in terms of the overall sample mean since all results for partial sample means can then be obtained as special cases.

The obvious expanded form of the term involving $\sigma^2_{(X)(Y)}$ is

$$\sigma^2_{(X)(Y)} \sum_{Z \subseteq Y} \frac{(-1)^q}{\prod\limits_{i \varepsilon Z} (\text{population range of } i)} \times \frac{1}{N_{X+(Y-Z)}}.$$

Let R be some fixed specified set of subscripts. In the completely expanded form of Ex^2 collect all terms for which σ^2 has the same X and for which $Y-Z = R$, i.e. vary Y and Z subject to the restriction that $Y-Z = R$. The sum of all such terms is

$$\frac{1}{N_{X+R}} \sum_{\substack{Z \subseteq Y \\ Y-Z=R}} \sigma^2_{(X)(Y)} \frac{(-1)^q}{\prod\limits_{i \varepsilon Z} (\text{population range of } i)} = \frac{1}{N_{X+R}} \Sigma_{(X)(R)},$$

by definition of the Σ's.

Hence

$$Ex^2 = \sum_{X,R} \left(\frac{\Sigma_{(X)(R)}}{N_{X+R}} \right) = \text{sum over all admissible pairs of } X \text{ and } R \text{ of } \frac{\Sigma_{(X)(R)}}{N_{X+R}}.$$

We notice that $N_{X+R} = \prod\limits_{i \varepsilon X+R} \alpha_i$, where α_i is the sample range of the index i in the complete sample, and $\alpha_i = 1$ when the index i ranges over the null set.

Let S_1 be the set of indices of any particular admissible partial sample mean. Then, relative to that mean, the number of different components whose type is specified by the set of indices $S_2 = X+R$ is $N = \prod\limits_{i \varepsilon S_2-S_1} \alpha_i$, and we see that an alternative statement of Theorem 1 is

$$Ex^2_{S_1} = \sum_{X,R} \left(\frac{\Sigma_{(X)(R)}}{\prod\limits_{i \varepsilon S_2-S_1} \alpha_i} \right).$$

There should be no confusion generated by the use of Σ for summation and for denoting 'cap sigmas'.

We have already noticed that for balanced samples in which inspection of a population value is attained through the use of a finite number of sampling stages only, the subscripts of a typical sample value and of a typical response correspond. The structure of the sample of observations is therefore identical with the structure of the population and hence there is a unique identity for the sample typical observation analogous to the population identity for the typical response. Sample components and degrees of freedom are defined analogously to the corresponding population quantities. Further, we define the sample analysis of variance (abbreviated

occasionally by ANOVA) according to type of component to be the tabular break-down of the total sum of squares and degrees of freedom of the sample observations corresponding line for line to the population analysis of variance. The possibility of such a decomposition for every balanced sample is ensured by the existence of the sample identity and hence by the validity for the sample structure of the assertions made earlier about population structures.

For each line of the sample analysis of variance we wish to include the expression for the expected value of the mean square of the line. We shall make a correspondence between the indices used to denote sample quantities and the subscripts of all the σ^2's and Σ's i.e. for the purpose of what follows we envisage identification of corresponding sample and population quantities by the same set of subscripts. Let S denote the complete set of subscripts used in writing all the σ^2's and Σ's, S_1 denote the set of subscripts associated with the leading mean of a particular line, and S_2 denote the set of subscripts associated with a particular σ^2 and the corresponding Σ. Further, let $S_1 = X_1 + Y_1$ and $S_2 = X_2 + Y_2$ where we denote by X's the respective sets of non-rightmost bracket subscripts and by Y's the sets of rightmost bracket subscripts. Let $W = Y_2 - Y_2 \cap S_1$, and let α_i and A_i denote respectively the sample (complete sample) and population ranges of the subscript i. Both these ranges are defined to be equal to one when the subscript i ranges over the null set.

We now state and prove the following theorem for arbitrary sample and population ranges.

Theorem 2 : *The expected value of the mean square of the line for which the set of subscripts for the leading mean is S_1 has the form*

$$\sum_{S_2 \subseteq S} R(S_1, S_2)\sigma^2_{S_2} = \sum_{S_2 \subseteq S} P(S_1, S_2) \Sigma_{S_2}$$

where the R's and P's are constants with values as follows

(i) $R(S_1, S_2) = P(S_2) = 0$ *if and only if S_2 does not contain S_1.*

(ii) *Whenever $S_1 \subseteq S_2$ then we may write*

$$P(S_1, S_2) = P(S_2) \text{ and } R(S_1, S_2) = P(S_2) \times Q(W), \text{ where } P(S_2) = \prod_{i \varepsilon S - S_2} \alpha_i$$

= *number of times any one component whose type is specified by S_2 enters into the complete sample used in the investigation and*

$$Q(W) = \prod_{i \varepsilon W} f_i^{\alpha_i} = \prod_{i \varepsilon W} \left(1 - \frac{sample\ range\ of\ index\ i}{population\ range\ of\ index\ i} \right).$$

Corollaries :   (1) *The mean squares of the population analysis of variance may be obtained, in either σ^2 or Σ forms, from the expected mean squares of the sample analysis of variance by replacing all sample quantities by the corresponding population quantities. Note that then the coefficients of all σ^2's vanish except the σ^2 for which $S_2^{\cdot} = S_1$.*

(2) *If for a particular σ^2 appearing in the general expression for the expected value of the mean square of a given line, at least one of the excess subscripts of a rightmost bracket, i.e. at least one $i \varepsilon W$, has its sample range equal to its population range, when the classification corresponding to the index is what is commonly called "fixed", then the contribution of that σ^2 vanishes from the expectation when the specific sample sizes are taken into consideration.*

(3) *The actual value of the finite correction factor corresponding to an index whose population range is infinite is equal to one under all circumstances.*

Proof of Theorem: The following proof makes use of Theorem 1 and its alternative statement and of the structure of lines in the ANOVA table.

Consideration of the scheme employed to denote sample observations and of assertion 5 shows that the expectation of the sum of squares for any line can be written as a product of the number of observations in the experiment and the expectation of a known linear function of squares of typical partial sample means. The form of both the function and the means is uniquely determined by the form of the leading mean of the line.

We begin by considering the contents of the expectation of the linear function with respect to a particular component of variation.

Suppose S_1 is not in S_2. Then the leading partial mean has at least one subscript, say i, appearing in its rightmost bracket but not appearing at all the component of variation. For every term containing the subscript i in the expanded form of the sample component defined by the leading mean, there is one of opposite sign and with subscripts identical with that of the other term but with i absent. Since the contents of the component of variation considered, and of the corresponding Σ, are identical in the expectation of squares of both these means it follows that the coefficient of these σ^2 and Σ is zero in the overall expectation of the line. This completes the proof of part (i) of the theorem.

It remains to determine the values of the R, P and Q for the case when $S_1 \subset S_2$. Though the direct derivation in terms of the σ^2's is instructive, we shall argue in terms of the simpler Σ forms.

Let Z be an arbitrary subset of Y_1 and let the number of subscripts in Z be denoted by q.

We first find the coefficient of Σ_{S_2} in the expected value of the sum of squares of the line for S_1. Consideration of the suggested form of the expected value of the sum of squares shows that coefficient to be:

$$\left(\prod_{i \varepsilon S} \alpha_i \right) \left(\frac{1}{\prod_{j \varepsilon S_2 - S_1} \alpha_j} \right) \left(\sum_{Z \subseteq Y_1} \frac{(-1)^q}{\prod_{k \varepsilon Z} \alpha_k} \right) = \left(\prod_{i \varepsilon S - S_2 + S_1} \alpha_i \right) \prod_{i \varepsilon Y_1} \alpha \left(1 - \frac{1}{\alpha_i} \right)$$

$$= \left(\prod_{i \varepsilon S - S_2 + X_1} \alpha_i \right) \left(\prod_{i \varepsilon Y_1} \alpha_i - 1 \right) = \left(\prod_{i \varepsilon S - S_2} \alpha_i \right) \times \left(\prod_{i \varepsilon X_1} \alpha_i \right) \prod_{i \varepsilon Y_1} (\alpha_i - 1).$$

Hence, the coefficient of Σ_{S_2} in the expected value of the mean square of the line for S_1 is $P(S_2) = \prod_{i \epsilon S - S_2} \alpha_i =$ number of times any one component whose type is specified by S_2 enters into the complete sample.

We now wish to find the coefficient of $\sigma_{S_2}^2 = \sigma_{(X_2)(Y_2)}^2$. From the definition of Σ's we know that $\sigma_{(X_2)(Y_2)}^2$ appears in those and only those Σ_{S_j}'s for which $S_2 - S_j = L_j \subset Y_2$. But in the present instance S_2 and all S_j's of relevance contain S_1 as a subset, so $S_2 - S_j$ contain no subscript of S_1, i.e. $L_j \cap S_1 = \theta$. Also, since $S_1 \subset S_2$ and because of the way in which permissible partial means are defined every subscript common to Y_2 and S_1 must appear in Y_1. Hence, the actual restriction on L_j is : $L_j \subset Y_2 - Y_1 \doteq W$.

The coefficient of Σ_{S_j} is $\prod_{i \epsilon S - S_j} \alpha_i = \prod_{i \epsilon (S - S_2) + (S_2 - S_j)} \alpha_i = \left(\prod_{i \epsilon S - S_2} \alpha_i \right) \left(\prod_{k \epsilon L_j} \alpha_k \right).$

Let the number of subscripts in L_j be q_j. Then the coefficient of $\sigma_{S_2}^2$ is seen to be

$$\left(\prod_{i \epsilon S - S_2} \alpha_i \right) \times \sum_{L_j \subseteq Y_2 - Y_1} \left(\prod_{k \epsilon L_j} \frac{(-1)^{q_j} \alpha_k}{A_k} \right)$$

$$= \left(\prod_{i \epsilon S - S_2} \alpha_i \right) \left(\prod_{k \epsilon Y_2 - Y_1} \left(1 - \frac{\alpha_k}{A_k} \right) \right)$$

$$= P(S_2) \times Q(W).$$

This completes the proof of the theorem. It should be noted that the σ^2 form of the theorem is essentially known and that equivalent forms are given by Bennett and Franklin (1954), and also Cornfield and Tukey (1956).

The merit of the present formulation lies in pointing out the simple form of the expected values of squares of sample means and how advantage of this may be taken when the structure of the lines of the sample analysis of variance table is defined precisely. Some tables exemplifying the application of the present theorem will be given later.

4. Randomized experiments

Up to this point we have considered the consequences of drawing pure samples, i.e. combinations of crossed and nested samples only. However, when as is done in actual designed experimental investigations, the physical act of randomization is also employed yet another type of sample is induced in some of the components. Thus, if a response, Y_{ik}, is determined by the treatment i and the unit k to which the treatment is applied, then in the sample the unit k can only be used once. A sample which is obtained by randomly applying the chosen treatments to the chosen experimental units, with r applications of each chosen treatment will then have a structure as indicated in Figure 2 below, where T and P correspond to the treatment and unit entities

respectively. A sample drawn according to such a scheme is said to be a simple fractionated sample.

entity P

entity T
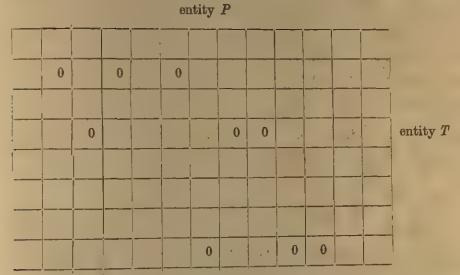
Fig. 2. A simple fractionated sample.

A sample of this type is induced in the treatment-experimental unit type of component, $(TP)_{ik}$. If the randomly chosen treatments are t in number out of T, and the randomly chosen units are rt out of P, then the expected value of the square of the overall induced sample mean, x, of the elements $(TP)_{ik}$ is given by

$$Ex^2 = \left[\left(1-\frac{1}{T} \right) - \frac{r}{P} \left(1-\frac{t}{T} \right) \right] \frac{\sigma^2_{(TP)}}{rt},$$

where

$$\sigma^2_{(TP)} = \frac{\sum\limits_{i,k} (TP)^2_{ik}}{(T-1)(P-1)}.$$

Again the form of this result has the immediate generalization in higher dimensions. Thus, if the treatments, denoted by A and B respectively, have a factorial structure then the generalization of the above formula is

$$Ex^2 = \left[\left(1-\frac{1}{A}\right) \left(1-\frac{1}{B}\right) - \frac{r}{P}\left(1-\frac{a}{A}\right)\left(1-\frac{b}{B}\right) \right] \frac{\sigma^2_{(ABP)}}{rab}.$$

Here a levels of A are chosen randomly out of A, and similarly and independently b levels of B out of B; also each chosen treatment combination is applied at random to exactly r experimental units, where altogether rab units are randomly chosen out of P and no unit can be used over in the experiment.

It should be noted that the second part of the correction factor bears a similarity to the type of correction factor obtained in the crossed case. Further, while the second part may for appropriate sample sizes become zero, this will never be the case with the first part of the correction factor.

We now formalize and prove our statements for the case of the two dimensional simple fractionated sample.

133

Consider a set of AP items with values denoted by P_{iu}, where $i = 1, 2, ..., A$; $u = 1, 2, ..., P$; and such that

$$\sum_{i=1}^{A} P_{iu} = \sum_{u=1}^{P} P_{iu} = 0.$$

Partition the P_{iu}'s into sets having like i's and choose at random a such sets. Within each chosen set select a random subset of size r but with the restriction that no chosen P_{iu}'s may have like values of u. Pictorially our sample will have the structure indicated in Figure 3. Introduce random variables α_i^{i*}, ρ_u^{i*f} with $i* = 1, 2, ..., a$; $i = 1, 2, ..., A$; $f = 1, 2, ..., r$; $u = 1, 2, ..., P$, and satisfying

$\alpha_i^{i*} = 1$ if the $i*$-th choice from among the i's selects all P_{iu}'s having subscript i,

$\alpha_i^{i*} = 0$ otherwise.

$\rho_u^{i*f} = 1$ if f-th selection within $i*$-th selected set corresponds to a P_{iu} value with second subscript $= u$,

$\rho_u^{i*f} = 0$ otherwise.

Then $\qquad E(\alpha_i^{i*})^2 = E(\alpha_i^{i*}) = \dfrac{1}{A}; \; E(\alpha_i^{i*} \alpha_{i'}^{i*'}) = \dfrac{1}{A(A-1)}$ for $i \neq i'$, $i* \neq i*'$

$$E(\rho_u^{i*f}) = \frac{1}{P}; \; E(\rho_u^{i*f} \rho_{u'}^{i*f'}) = \frac{1}{P(P-1)} \quad \begin{matrix} u \neq u' \\ f \neq f' \end{matrix}$$

$$E(\rho_u^{i*f} \rho_{u'}^{i*'f'}) = \frac{1}{P(P-1)}; i* \neq i*', u \neq u', \text{ and all values of } f \text{ and } f'.$$

Denote the $(i*f)$-th selected value by x_{i*f}. Then

$$x_{i*f} = \sum_{iu} \alpha_i^{i*} \rho_u^{i*f} P_{iu}.$$

Theorem 3 : *The expectation of the square of a sample mean selected by the procedure specified above is*

$$Ex^2 = E\left(\frac{1}{ar} \sum_{i*f} x_{i*f}\right)^2 = \frac{\sigma_{AP}^2}{ar}\left(\left(1 - \frac{1}{A}\right) - \frac{r}{P}\left(1 - \frac{a}{A}\right)\right).$$

Proof: $\quad E\left(\dfrac{1}{ar} \sum_{i*f} x_{i*f}\right)^2 = E \sum \dfrac{1}{a^2} \sum_{i*,i} (\alpha_i^{i*}) \dfrac{1}{r^2} \sum_{f,f'} \sum_{u,u'} \rho_u^{i*f} \rho_{u'}^{i*f'} P_{iu}P_{iu'}$

$$+ E\frac{1}{a^2} \sum_{i* \neq i*'} \sum_{i \neq i'} (\alpha_i^{i*} \alpha_{i'}^{i*'}) \frac{1}{r^2} \sum_{f,f'} \sum_{u \neq u'} \rho_u^{i*f} \rho_{u'}^{i*'f'} P_{iu} P_{i'u'}.$$

Now, $\quad E\left(\dfrac{1}{r^2} \sum_{f,f'} \sum_{u,u'} \rho_u^{i*f} \rho_{u'}^{i*f'} P_{iu} P_{iu'}\right) = E\left(\dfrac{1}{r} \sum_{f,u} \rho_u^{i*f} P_{iu}\right)^2 = \dfrac{\sum_u P_{iu}^2}{(P-1)r}\left(1 - \dfrac{r}{P}\right).$

Also, $\quad E\dfrac{1}{r^2} \sum_{f,f'} \sum_{u \neq u'} \rho_u^{i*f} \rho_{u'}^{i*'f'} P_{iu} P_{i'u'}$ with $i \neq i'$, $i* \neq i*'$

$$= \frac{1}{P(P-1)} \sum_u P_{iu} P_{i'u}.$$

Hence
$$E\left(\frac{1}{ar}\sum_{i \neq f} x_{i \bullet f}\right)^2 = \frac{1}{a^2}\sum_{i \bullet, i} E(\alpha_i^{i \bullet})\frac{\sum_u P_{iu}^2}{(P-1)\,r}\left(1-\frac{r}{P}\right)$$

$$+\frac{1}{a^2}\sum_{i \bullet \neq i \bullet'}\sum_{i \neq i'} E(\alpha_i^{i \bullet}\,\alpha_{i'}^{i \bullet'})\left(-\frac{1}{P(P-1)}\sum_u P_{iu}\,P_{i'u}\right)$$

$$=\frac{1}{aA}\frac{\sum_{iu}P_{iu}^2}{(P-1)r}\left(1-\frac{r}{P}\right)+\frac{a-1}{aA(A-1)}\frac{\sum_{iu}P_{iu}^2}{P(P-1)}$$

$$=\frac{\sigma_{AP}^2}{ar}\left(1-\frac{1}{A}\right)\left(1-\frac{r}{P}\right)+\frac{r(a-1)}{AP}\frac{\sigma_{AP}^2}{ar}$$

$$=\frac{\sigma_{AP}^2}{ar}\left(\left(1-\frac{1}{A}\right)-\frac{r}{P}\left(1-\frac{a}{A}\right)\right). \quad \text{q.e.d.}$$

As mentioned before the generalization which suggests itself by this two-dimensional case holds also when the number of dimensions is $R+1$, $R > 1$. Proof of this may be developed along the above lines and making use of the method of induction. The details are given by Zyskind (1958).

Again, as in the case of balanced, crossed and nested samples, the correction factors involved in simple fractionated samples involve only letters corresponding to subscripts of the rightmost bracket of the component in question. For example, in a randomized block the plots, P, are nested in the blocks, B, so letters of the rightmost bracket of the interaction of the treatments with plots within blocks are T and P. The expected value of the square of the overall mean induced in this interaction is

$$\left[\left(1-\frac{1}{T}\right)-\frac{r}{P}\left(1-\frac{t}{T}\right)\right]\frac{\sigma_{(B)(PT)}^2}{rbt},$$

where b is the number of randomly chosen blocks out of B and

$$\sigma_{(B)(PT)}^2 = \frac{\sum_{i,j,k}(PT)_{k(jk)}^2}{B(P-1)(T-1)}.$$

We now pause to illustrate by means of the randomized block design in some degree of detail the nature of our results this far.

As stated before the population structure for the randomized block is

$$(B:P)(T) \quad \text{or} \quad (i:j)(k).$$

The population identity for the typical response Y_{ijk} therefore is

$$Y_{ijk} = Y+(Y_i-Y)+(Y_k-Y)+(Y_{ik}-Y_i-Y_k+Y_{ik})$$
$$+(Y_{ij}-Y_i)+(Y_{i(jk)}-Y_{ij}-Y_{ik}+Y_i)$$
$$=\mu+B_i+T_k+(BT)_{ik}+P_{i(j)}+(TP)_{i(jk)}.$$

The experimental procedure consists of choosing randomly b blocks out of B, and in each chosen block rt experimental units out of P; of selecting t treatments out of T; and within each chosen block applying at random each of the t selected treatments

to exactly r experimental units. Denote by x_{i*k*f} the observed value associated with the f-th selected repetition of the k^*-th chosen treatment in the i^*-th selected block. The sample observational structure relative to this notation therefore is $(i^*k^*:f)$. This implies the sample observational identity

$$x_{i*k*f}=x+(x_{i*}-x)+(x_{k*}-x)+(x_{i*k*}-x_{i*}-x_{k*}+x)+(x_{i*k*(f)}-x_{i*k*}),$$

and the corresponding sample analysis of variance.

In order to link the observed values to the population quantities we introduce sets of sample and design random variables as follows:

$$(\alpha_i^{i*}),\ (\mu_{i*j}^{i*j*}),\ (\beta_k^{k*}),\ (\lambda_{i*j*}^{i*k*}),\ \text{and}\ (\rho_{i*j}^{i*k*f}),\ \text{where}\ \rho_{i*j}^{i*k*f}=\sum_{j*=1}^{rt}\mu_{i*j}^{i*j*}\ \lambda_{i*j*}^{i*k*f}\ ;$$

where again the meaning of the separate variables should be clear from the context, and their elementary properties can be easily derived. The statistical model is

$$x_{i*k*f}=\sum_{ijj*k}\alpha_i^{i*}\ \beta_k^{k*}\ \mu_{i*j}^{i*j*}\ \lambda_{i*j*}^{i*k*f}\ Y_{ijk}=\sum_{ijk}\alpha_i^{i*}\ \beta_k^{k*}\ \rho_{i*j}^{i*k*f}\ Y_{ijk}$$

$$=\mu+\sum_i\alpha_i^{i*}\ B_i+\sum_k\beta_k^{k*}\ T_k+\sum_{ik}\alpha_i^{i*}\ \beta_k^{k*}\ (BT)_{ik}$$

$$+\sum_{ij}\alpha_i^{i*}\ \rho_{i*j}^{i*k*f}\ P_{i(j)}+\sum_{ijk}\alpha_i^{i*}\ \rho_{i*j}^{i*k*f}\ \beta_k^{k*}\ (TP)_{i(jk)}.$$

By making use directly of the properties of the random variables involved or otherwise by observing the types of samples that are induced in the various types of population components we can easily see that

$$Ex^2=\mu^2+\left(1-\frac{b}{B}\right)\frac{\sigma_{(B)}^2}{b}+\left(1-\frac{t}{T}\right)\frac{\sigma_{(T)}^2}{t}+\left(1-\frac{b}{B}\right)\left(1-\frac{t}{T}\right)\frac{\sigma_{(BT)}^2}{bt}$$

$$+\left(1-\frac{rt}{P}\right)\frac{\sigma_{(B)(P)}^2}{rtb}+\left(\left(1-\frac{1}{T}\right)-\frac{r}{P}\left(1-\frac{t}{T}\right)\right)\frac{\sigma_{(B)(PT)}^2}{rbt}.$$

On regrouping terms we obtain

$$Ex^2=\left(\mu^2-\frac{1}{B}\ \sigma_{(B)}^2-\frac{1}{T}\ \sigma_{(T)}^2+\frac{1}{BT}\ \sigma_{(BT)}^2\right)$$

$$+\frac{1}{b}\left(\sigma_{(B)}^2-\frac{1}{T}\sigma_{(BT)}^2-\frac{1}{P}\sigma_{(B)(P)}^2+\frac{1}{PT}\ \sigma_{(B)(PT)}^2\right)$$

$$+\frac{1}{t}\left(\sigma_{(T)}^2-\frac{1}{B}\sigma_{(BT)}^2\right)+\frac{1}{bt}\left(\sigma_{(BT)}^2-\frac{1}{P}\sigma_{(B)(PT)}^2\right)$$

$$+\frac{1}{rtb}\left(\sigma_{(B)(P)}^2-\frac{1}{T}\sigma_{(B)(PT)}^2\right)+\frac{1}{rbt}\ \sigma_{(B)(PT)}^2.$$

On insertion of Σ terms the above expression assumes the simple form :

$$Ex^2=\Sigma_\theta+\frac{1}{b}\ \Sigma_{(B)}+\frac{1}{t}\ \Sigma_{(T)}+\frac{1}{bt}\Sigma_{(BT)}+\frac{1}{rbt}\left(\Sigma_{(B)(P)}+\Sigma_{(B)(PT)}\right).$$

As might have been anticipated, the expression for Ex^2 in terms of the Σ's retains the form it had in the case of ordinary balanced samples only. The additional operation of random confounding has not altered the form of the Σ expansion for Ex^2. As before the coefficient of each Σ is the reciprocal of the number of possibly different values of the type of component in question entering into the formation of the overall sample mean.

The applicability of the form of the above expression is very general indeed. It has been verified to hold in the generalized case of the completely randomized, randomized block, latin square, and various modifications of the split-plot designs. It has also been shown to hold for the general case of the balanced incomplete block design and for a fair number of detailed situations in which treatments are subject to error.

Expected values for partial observational means can easily be obtained from the above when the simple expansions for Ex^2 are valid by simple substitution. Thus, in the case of the randomized block if x_{i*} is the observed mean of the i*-th chosen block, then Ex_{i*}^2 is obtained from the expression for Ex^2 by substituting the value $b = 1$. Thus,

$$Ex_{i*}^2 = \Sigma_{(\theta)} + \Sigma_{(B)} + \frac{\Sigma_{(T)}}{t} + \frac{\Sigma_{(BT)}}{t} + \frac{\Sigma_{(P)(T)}}{rt} + \frac{\Sigma_{(B)(PT)}}{rt}.$$

The actual simplicity of the above form may perhaps better be put in evidence by pointing out its complete analogy with the σ^2 forms obtained in taking expectations of squares of partial means in a two-way table, where all the components, apart from the overall mean, are completely random and uncorrelated.

Thus, consider the model

$$y_{ij} = \mu + a_i + b_j + (ab)_{ij} \qquad i = 1, 2, ..., a; \qquad j = 1, 2, ..., b$$

where μ is a constant, the a_i's are a random sample of size a and b_j's a sample of size b, the $(ab)_{ij}$'s a sample of size ab, all samples are independent, and from infinite populations with means zero and finite variances. Denote μ^2 by $\sigma_\mu^2 = \sigma_\theta^2$.

$$\text{var} \ (a_i) = \sigma_a^2, \quad \text{var} \ (b_j) = \sigma_b^2, \quad \text{var} \ (ab)_{ij} = \sigma_{ab}^2.$$

Then

$$Ey_{ij}^2 = \sigma_\theta^2 + \sigma_a^2 + \sigma_b^2 + \sigma_{ab}^2,$$

$$E(y_i^2) = E\left(\frac{1}{b} \sum_{j=1}^{b} y_{ij}\right)^2 = \sigma_\theta^2 + \sigma_a^2 + \frac{\sigma_b^2}{b} + \frac{\sigma_{ab}^2}{b},$$

$$E(y_j^2) = E\left(\frac{1}{a} \sum_{i=1}^{a} y_{ij}\right)^2 = \sigma_\theta^2 + \frac{\sigma_a^2}{a} + \sigma_b^2 + \frac{\sigma_{ab}^2}{a},$$

$$E(y^2) = E\left(\frac{1}{ab} \sum_{ij} y_{ij}\right)^2 = \sigma_\theta^2 + \frac{\sigma_a^2}{a} + \frac{\sigma_b^2}{b} + \frac{\sigma_{ab}^2}{ab},$$

i.e. the contents of σ_i^2, $i = \theta, a, b, ab$; in the expectation of the square of a particular sample mean is equal to σ_i^2 divided by the number of possibly different i elements entering into the particular sample mean. With proper choice of sample indices the analogy can also be carried forcefully to the corresponding analysis of variance tables of expected mean squares, since for orthogonal sample analyses of variance each such expected mean square can be written out simply and uniquely as a known linear function of expected values of squares of partial observational means.

We now formalize our results. Let $\Sigma_{(X)(R)}$ denote the Σ corresponding to the type of population component for which the set of letters belonging to the rightmost bracket is R and for which X is the set of letters not belonging to the rightmost bracket. Let $N_{(X)(R)}$ denote the number of values of component type $(X)(R)$ entering into the formation of the complete experimental sample. Then Ex^2 is said to admit the standard Σ expansion if

$$Ex^2 = \sum_{X,R} \frac{\Sigma_{(X)(R)}}{N_{(X)(R)}}.$$

Denote the subscripts of some particular admissible sample partial mean by S_1^* and by $N_{(X)(R), S_1^*}$ the number of possibly different values of components of type $(X)(R)$ entering into the formation of $x_{S_1^*}$. Then if Ex^2 admits the standard Σ expansion it follows that

$$Ex^2_{S_1^*} = \sum_{X,R} \frac{\Sigma_{(X)(R)}}{N_{(X)(R), S_1^*}}.$$

A sufficient condition for the standard Σ form of Ex^2 to hold is that all of the balanced samples involved be of the cross, nested, or simple fractionated type. A more relaxed sufficient condition may be obtained by realizing that coefficients, preceding σ^2's which arise in the above situations satisfy also the condition exemplified by the following illustration.

The full term involving $\sigma^2_{(B)(PT)}$ in the σ^2 expression for Ex^2 in the case of the randomized block is

$$\left(\left(1 - \frac{1}{T}\right) - \frac{r}{P}\left(1 - \frac{t}{T}\right) \right) \frac{\sigma^2_{(B)(PT)}}{rbt}.$$

It can also be written as

$$\frac{1}{rbt} \times \frac{\sigma^2_{(B)(PT)}}{1} - \frac{1}{rbt} \times \frac{\sigma^2_{(B)(PT)}}{T} - \frac{1}{bt} \times \frac{\sigma^2_{(B)(PT)}}{P} + \frac{1}{b} \frac{\sigma^2_{(B)(PT)}}{PT}.$$

If we denote by Z the variable set of letters in the denominator of the expressions of which $\sigma^2_{(B)(PT)}$ is the numerator above, and if we denote by $N_{(B)(PT-Z)}$ the number of values of components of type $(B)(PT-Z)$ entering into the formation of the observed experimental sample mean which induced the sample in question in the component $(B)(PT)$, then we see that every term above is of the form

$$\frac{(-1)^q}{N_{(B)(PT-Z)}} \times \frac{\sigma^2_{(B)(PT)}}{Z},$$

where q is the number of letters in Z. The nature of the above form is one which obtains under a very wide range of experimental conditions. Its applicability forms a sufficient condition for standard Σ expansions of the expected values of squares of partial sample means.

Formalizing then we now establish the following statement.

Theorem 4 : *Consider the situations where :*

(1) *The sample means induced by the experimental procedure in the different types of population components are pairwise uncorrelated. Hence, the expected value of the square of an admissible observed sample mean is equal to the sum of the expected values of squares of sample means from the individual types of components.*

(2) *The induced samples are such that for each type of component the expected value of the square of an induced mean is equal to the corresponding component of variation divided by the number of values in the mean and multiplied by a product of correction factors specified as follows. Each correction factor corresponds to one of a number of disjoint groups of indices of the rightmost bracket, where the totality of the groups exhausts the rightmost bracket, and is of a form arising from either balanced nested, crossed, or simple fractionated samples.*

Alternatively, consider the situation where :

(3) *The term involving* $\sigma^2_{(X)(Y)}$ *in the* σ^2 *form of the expansion of* Ex^2 *is*

$$\sigma^2_{(X)(Y)} \sum_{Z \subseteq Y} \frac{(-1)^q}{\prod\limits_{i \in Z} (range \ of \ i)} \times \frac{1}{N_{(X)(Y-Z)}}.$$

Inspection reveals that (1) and (2) imply (3). The converse, however, does not necessarily hold. The argument on page 129 now shows that whenever for each admissible $\sigma^2_{(X)(Y)}$ the corresponding term is given by (3) then

$$Ex^2 = \sum_{X,R} \left(\frac{\Sigma_{(X)(R)}}{N_{(X)(R)}} \right).$$

Denote the subscripts of some admissible sample partial mean by S_1^* and by $N_{(X)(R), S_1^*}$, $S_1^* = N_{S_2}$, $S_1^* =$ the number of possibly different components of type S_2 entering into the formation of $x_{S_1^*}$. Then it follows that

$$Ex^2_{S_1^*} = \sum_{X,R} \left(\frac{\Sigma_{(X)(R)}}{N_{(X)(R), S_1^*}} \right). \quad \text{q.e.d.}$$

With regard to tables of the sample analysis of variance we remark that their subdivision into individual lines is completely determined by the internal structure imposed on the sample set of indices. Each admissible sample mean becomes the forming or leading term of one of the lines of the table, and with balance the total sum of squares decomposes additively into the sum of the sums of squares belonging to the individual lines.

139

In what follows an attempt is made to derive sufficient conditions for the expected values of the mean squares of the lines of analyses of variance tables to be "simple", in terms of the Σ's. The conditions are similar to the ones needed with cases where cross and nested samples only are involved. However, because with randomization the correspondence between sample and population subscript is no longer one-one throughout, the additional concept of ambivalence, attempting to emphasize the possible relationships, has been introduced. It is believed that as more understanding of the situations is gained, a simpler descriptive scheme will be formulated. Work on such a scheme is now in progress.

Definition : A sample subscript i^* and a population subscript j are said to be in the relation of complete ambivalence if summation in the observed sample values over the range of i^* always implies that the size of the induced sample in every one of the population components involving j is multiplied by the sample range of i^*.

As a concrete example consider again the case of the randomized block design.

The population decomposition of Y_{ijk} into population components and the statistical model for x_{i*k*f} are as on pages 135 and 136. It is easy to see that any summation, in addition to previous summations over i^* and / or f, of the x_{i*k*f}'s over the range t of the sample index k^* causes the induced sample size of every population component involving the subscript k to be t times as large. The same happens for every population component involving the subscript j. Thus, in a certain sense, the sample subscript k^* corresponds completely to the population subscripts j and k. We say that k^* is in complete ambivalence relation with j and k, or equivalently with the corresponding population entities P and T.

Definition : The full sets of sample and population subscripts are said to be in the relation of complete ambivalence if whenever summation over the range of any one of the sample indices is effected, then independently of the state of the other indices the size of the induced sample in every one of the types of population components either remains unchanged or is multiplied by the range of the sample index in question.

Thus, referring again to page 136 we verify that always summation in the sample observations over the range of k^* induces the multiplicative factor t in every population component involving k or j and induces no change in the sample size of any of the other population components. Similarly, summation over the sample range, r, of the index f involves multiplication by r of the induced sample size of every population component having j for a subscript, and involves no change in the induced sample size of the other population components. Finally, summations in the observed values over the sample range of i^* induces in a similar way a change in all population components having i or j for subscripts and induces no change in any of the other components. Thus, in this example the full sets of sample and population subscripts are said to be in the relation of complete ambivalence.

We mention in passing that in the case of the latin square the reasonable full set of sample subscripts is not in complete ambivalence relation with the full set of population subscripts. Hence, the theorem below on the expected values of mean squares cannot be applied as it stands to the latin square case. However, the expected values of squares of admissible sample means still admit the standard Σ expansion. Hence, expressions for the expected mean squares can be easily obtained for the latin square by direct manipulation. In like manner, if one does not wish to make use of the notion of complete ambivalence in the concrete examples encountered in practice, one can still obtain easily all of the expected mean squares by computing simple linear combinations of the relevant Σ forms.

Consider the set of experiments for which the expected value of the square of the overall sample mean admits a standard Σ expansion. Suppose further that the set of sample subscripts used is in complete ambivalence relation with the set of population subscripts. Let $x_{S_1^*}$ be the leading term of a particular line of the sample analysis of variance table and Y_1^* be the set of rightmost bracket subscripts of S_1^*. Let S_2 denote the set of subscripts belonging to a typical population component.

Theorem 5 : *The expected value of the mean square whose leading term is* $x_{S_1}^*$ *is* $\sum_{S_2} P(S_2) \, \Sigma_{S_2}$; *where* $P(S_2)$ *has values as follows* : $P(S_2) = 0$ *if and only if at least one of the subscripts of* Y_1^* *is not in complete ambivalence relation with any of the subscripts of* S_2; *and otherwise* $P(S_2) =$ *number of times a value of typical component of* S_2 *enters into the formation of the complete observed sample.*

Proof : Let S_2 be the set of subscripts defining any one particular type of component. Then, by the argument of the theorem on expected mean squares of Section 3 the coefficient of Σ_{S_2} in the expected value of the mean square is zero if at least one of the subscripts of Y_1^* is not in ambivalence relation with any of the subscripts of S_2. Again, if every subscript of Y_1^* is in ambivalence relation with at least one of the subscripts of S_2, then the coefficient of Σ_{S_2} in the expected value of the sum of squares corresponding to S_1^* is

$$\left(\underset{i^*\varepsilon S^*}{\pi} \alpha_{i^*} \right)\left(\frac{1}{\underset{j^*\varepsilon Z_{S^*-S_1^*,S_2}}{\pi} \alpha_{j^*}} \left(\sum_{Z^* \subseteq Y_1^*} \frac{(-1)^q}{\underset{m^*\varepsilon Z^*}{\pi} \alpha_{m^*}} \right) \right)$$

$$= \left(\underset{i^*\varepsilon S^*}{\pi} \alpha_{i^*} \right) \left(\frac{1}{\underset{j^*\varepsilon Z_{S^*-S_1^*,S_2}}{\pi} \alpha_{j^*}} \right)\left(\frac{1}{\underset{m^*\varepsilon Y_1^*}{\pi} \alpha_{m^*}} \right)\left(\underset{m^*\varepsilon Y_1^*}{\pi} (\alpha_{m^*}-1) \right)$$

$$= \left(\underset{i^*\varepsilon S^*-S_1^*}{\pi} \alpha_{i^*} \right)\left(\frac{1}{\underset{j^*\varepsilon Z_{S^*-S_1^*,S_2}}{\pi} \alpha_{j^*}} \right) \times \text{(D.F. of } S_1^*).$$

Hence the coefficient of ΣS_2 in the expected value of the mean square corresponding to $x_{S^\bullet}$ is equal to the product of sample ranges of all indices in $S^* - S_1^\bullet$ which are not in ambivalence relation with any of the subscripts of S_2. However, because of the way in which admissible means are constructed every subscript of $S_1^\bullet$ is in ambivalence relation with at least one of the subscripts of S_2. Hence, the required number is equal to the product of sample ranges of indices of S^* which are not in ambivalence relation with any of the subscripts of S_2; and so is equal to the number of times a typical selected value of the component of types S_2 enters into the formation of the complete observed sample.   q.e.d.

Corollary (i) :   *Whenever ΣS_2 appears with non-zero coefficient in the expected value of a mean square, then so does ΣS_3 for every $S_3 \supseteq S_2$.*

Corollary (ii) :   *The expected value of a mean square with rightmost bracket of the leading term $Y_1^\bullet$ is equal to the common part (i.e. intersection) of the expected mean squares of any two lines for which the respective sets of rightmost bracket subscripts $U_1^\bullet$ and $W_1^\bullet$ are such that*

$$U_1^\bullet + W_1^\bullet = Y_1^\bullet.$$

It should be noted that whenever a sample subscript is in ambivalence relation with a given population subscript, say K, then it is also in ambivalence relation with every subscript nested in K. For this reason we shall indicate explicitly the relationship with K only. Further, if the index K corresponds to the population entity, say T, then we shall also say that the sample index is in ambivalence relation with T.

In the case of a fixed set of treatment effects, the following simple result is of value.

Let each treatment enter into the experiment K times. Denote by T^* the mean square due to treatments, and by σ_T^2 the treatment component of variation. Then the average variance of the natural unbiassed estimates of the treatment differences is given by

$$\frac{2}{K} \, (E(T^*) - K \sigma_T^2).$$

It should be noted that the theorem of the present section on expected mean squares contains the better known one involving balanced crossed and nested samples only as a special simple particular case which is also simple to state. In the appendix we exemplify with three analyses of variance, two involving the relations of crossing and nesting only and the last involving also the relation of simple fractionation in a situation in which the treatment levels are subject to error. A number of other specific tables may be found (Wilk, 1955), (Zyskind, 1958), (Zyskind and Kempthorne, 1960).

Appendix

SOME EXAMPLES

Example 1 : Envisage an investigation of the fertility structure with respect to some uniform treatment of a specified agricultural land area. Suppose the initially postulated structure is

$$(S:(R) \ (C:L)).$$

Suppose that the sampling operation consists of randomly selecting s sources out of S, in each selected source choosing randomly r rows out of the R and c columns out of the C, and then within each chosen column choosing l columns out of L. We exhibit the complete analysis of variance in Table 1.

Example 2 : Four entities are involved in the structure. Denote them by P, Q, S, and R. The set of relations is : Q is nested in S, and R is nested in SP combinations. Symbolically :

$$(S:Q) \ (P) \quad \text{and} \quad (SP:R).$$

Using these letters for subscripts we see, that the population identity for the typical response is

$$Y_{SPQR} = Y + (Y_S - Y) + (Y_P - Y) + (Y_{SP} - Y_S - Y_P + Y)$$
$$+ (Y_{S(Q)} - Y_S) + (Y_{S(QP)} - Y_{SQ} - Y_{SP} + Y_S)$$
$$+ (Y_{SP(R)} - Y_{SP}) + (Y_{SP(QR)} - Y_{SPQ} - Y_{SPR} + Y_{SP}).$$

Denote the respective population ranges of the subscripts by S, P, Q and R. Then, using the obvious notational correspondence we verify that the Σ's are as given on page 122.

Envisage a sampling investigation conducted on the above structure. Denote the sample ranges by small letters corresponding to the capital letters used for the population ranges. We exhibit the resulting analysis of variance in Table 2.

Example 3 : We now consider a specialized procedure in an experiment in generalized randomized blocks when the application of intended amounts of treatments may be subject to error. The errors in the treatments are considered as treatment sublevels nested within treatment levels. A more detailed treatment of the following may be found in Zyskind and Kempthorne (1960).

The population structure is envisaged to be as follows. The experimental material has a nested structure : that of the blocks, B, nesting in them the ultimate experimental units, P (for plots). Again, the main levels of the treatment entity, G, nest within them the treatment sublevels, g. The population structure may, therefore, be symbolically represented by

$$(B:P) \ (G:g).$$

We denote the respective indices ranging over the elements of these entities by i, j, k, and m. Also, except for m, we denote the ranges of these indices by the capital letters denoting the entities themselves; thus $i = 1, 2, .., B; j = 1, 2, .., P; k = 1, 2, ..., G$; also $m = 1, 2, .., M$. An alternative representation of the population structure therefore is

$$(i:j) \ (k:m).$$

The experimental procedure consists of choosing firstly, b blocks at random out of B and g treatment levels (or treatments) out of G. Further, suppose that the b chosen blocks are arranged at random into w groups of q blocks each ($wq = b$). A treatment level is applied simultaneously to all the rq plots within a group of blocks which receive it. Hence, treatment sublevels are randomly confounded with blocks since the blocks are arranged into groups at random.

We denote by $x_{i*m*k*f}$ the observed sample value in the m^*-th chosen block in the i^*-th selected group, of the f-th repetition of the k^*-th chosen treatment. The sets of random variables we introduce are :

$$(B_i^{i*m*}), (G_k^{k*}), (g_{k*m}^{k*i*}), (\rho_{i*m*j}^{i*m*k*f}).$$

143

The meaning and elementary properties of these variables should be apparent from the content. The statistical model is

$$x_{i \bullet m \bullet k \bullet f} = \sum_{ijkm} B_i^{i \bullet m \bullet} G_k^{k \bullet} g_{k \bullet m}^{k \bullet i \bullet} \rho_{i \bullet m \bullet j}^{i \bullet m \bullet k \bullet f} Y_{ijkm}$$

$$= \mu + \sum_i B_i^{i \bullet m \bullet} B_i + \sum_k G_k^{k \bullet} G_k + \sum_{ik} B_i^{i \bullet m \bullet} G_k^{k \bullet} (BG)_{ik}$$

$$+ \sum_{ij} B_i^{i \bullet m \bullet} \rho_{i \bullet m \bullet j}^{i \bullet m \bullet k \bullet f} P_{i(j)} + \sum_{ijk} B_i^{i \bullet m} G_k^{k \bullet} \rho_{i \bullet m \bullet j}^{i \bullet m \bullet k \bullet f} (PG)_{i(jk)}$$

$$+ \sum_{km} G_k^{k \bullet} g_{k \bullet m}^{k \bullet i \bullet} g_{k(m)} + \sum_{kim} G_k^{k \bullet} B_i^{i \bullet m \bullet} g_{k \bullet m}^{k \bullet i \bullet} (Bg)_{k(im)}$$

$$+ \sum_{ikfm} B_i^{i \bullet m \bullet} G_k^{k \bullet} \rho_{i \bullet m \bullet j}^{i \bullet m \bullet k \bullet f} g_{k \bullet m}^{k \bullet i \bullet} (Pg)_{ik(jm)}.$$

Using the properties of the random variables or by inspecting the types of samples induced it can be seen that

$$Ex^2 = \mu^2 + \left(1 - \frac{b}{B}\right) \frac{\sigma_{(B)}^2}{b} + \left(1 - \frac{g}{G}\right) \frac{\sigma_{(G)}^2}{g}$$

$$+ \left(1 - \frac{b}{B}\right)\left(1 - \frac{g}{G}\right) \frac{\sigma_{(BG)}^2}{bg} + \left(1 - \frac{rg}{P}\right) \frac{\sigma_{(B)(P)}^2}{rg w q}$$

$$+ \left(\left(1 - \frac{1}{G}\right) - \frac{r}{P}\left(1 - \frac{g}{G}\right)\right) \frac{\sigma_{(B)(PG)}^2}{rg w q} + \left(1 - \frac{w}{M}\right) \frac{\sigma_{(G)(g)}^2}{wg}$$

$$+ \left(\left(1 - \frac{1}{M}\right) - \frac{q}{B}\left(1 - \frac{w}{M}\right)\right) \frac{\sigma_{(G)(Bg)}^2}{q w g}$$

$$+ \left(\left(1 - \frac{1}{M}\right) - \frac{r}{P}\left(1 - \frac{1}{M}\right)\right) \frac{\sigma_{(BG)(Pg)}^2}{r q w g}$$

$$= \Sigma_{(\rho)} + \frac{1}{b}\Sigma_{(B)} + \frac{1}{g}\Sigma_{(G)} + \frac{1}{bg}\Sigma_{(BG)} + \frac{1}{brg}\Sigma_{(B)(P)}$$

$$+ \frac{1}{brg}\Sigma_{(B)(PG)} + \frac{1}{wg}\Sigma_{(G)(g)} + \frac{1}{qwg}\Sigma_{(G)(Bg)}$$

$$+ \frac{1}{rqwg}\Sigma_{(BG)(Pg)}.$$

Further, since the ambivalence relation between the set of sample and population subscripts is complete, all of the forms for expected values of mean squares exhibited in Table 3 follow as a direct consequence of the Theorem 5. Alternatively, Table 3 may be easily obtained by direct manipulation of expected values of squares of the appropriate sample means. In Table 3 Σ' stands for the sum of all Σ's whose sets of subscripts involve P. We note that with a notation which ignored the fact that groups of blocks are treated alike with respect to the sublevels, the complete forms of the ensuing analysis of variance table would no longer be "simple".

The more frequently treated case in which there are no treatment errors may be easily obtained from the present one by deleting from consideration all components involving treatment sublevels, i.e. all components involving the subscript g.

TABLE 1. THE SAMPLE ANALYSIS OF VARIANCE FOR THE STRUCTURE $(S:(R)(C:L))$

source of variation and leading mean	d.f.	e.m.s.
mean $X_{(\phi)}$	1	$lrcs\,\Sigma_{(\theta)} + lrc\,\Sigma_{(S)} + lc\,\Sigma_{(S)(R)} + lr\,\Sigma_{(S)(C)} + l\,\Sigma_{(S)(RC)}$ $+\, r\,\Sigma_{(S)(C)(L)} + \Sigma_{(S)(C)(RL)}$ $= lrcs\sigma^2_{(\theta)} + lrc\left(1-\dfrac{s}{S}\right)\sigma^2_{(S)} + lc\left(1-\dfrac{r}{R}\right)\sigma^2_{(S)(R)}$ $+\, lr\left(1-\dfrac{c}{C}\right)\sigma^2_{(S)(C)} + l\left(1-\dfrac{r}{R}\right)\left(1-\dfrac{c}{C}\right)\sigma^2_{(S)(RC)}$ $+\, r\left(1-\dfrac{l}{L}\right)\sigma^2_{(S)(C)(L)} + \left(1-\dfrac{r}{R}\right)\left(1-\dfrac{l}{L}\right)\sigma^2_{(S)(C)(RL)}$
sources $X_{(S)}$	$s-1$	$lrc\,\Sigma_{(S)} + lc\,\Sigma_{(S)(R)} + lr\,\Sigma_{(S)(C)} + l\,\Sigma_{(S)(RC)} + r\,\Sigma_{(S)(C)(L)}$ $+\,\Sigma_{(S)(C)(RL)}$ $= lrc\,\sigma^2_{(S)} + lc\left(1-\dfrac{r}{R}\right)\sigma^2_{(S)(R)} + lr\left(1-\dfrac{c}{C}\right)\sigma^2_{(S)(C)}$ $+\, l\left(1-\dfrac{r}{R}\right)\left(1-\dfrac{c}{C}\right)\sigma^2_{(S)(RC)} + r\left(1-\dfrac{l}{L}\right)\sigma^2_{(S)(C)(L)}$ $+\left(1-\dfrac{r}{R}\right)\left(1-\dfrac{l}{L}\right)\sigma^2_{(S)(C)(RL)}$
rows within sources $X_{(S)(R)}$	$s(r-1)$	$lc\,\Sigma_{(S)(R)} + l\,\Sigma_{(S)(RC)} + \Sigma_{(S)(C)(RL)}$ $= lc\,\sigma^2_{(S)(R)} + l\left(1-\dfrac{c}{C}\right)\sigma^2_{(S)(RC)} + \left(1-\dfrac{l}{L}\right)\sigma^2_{(S)(C)(RL)}$
columns within sources $X_{(S)(C)}$	$s(c-1)$	$lr\,\Sigma_{(S)(C)} + l\,\Sigma_{(S)(RC)} + r\,\Sigma_{(S)(C)(L)} + \Sigma_{(S)(C)(RL)}$ $= lc\,\sigma^2_{(S)(C)} + l\left(1-\dfrac{r}{R}\right)\sigma^2_{(S)(RC)} + r\left(1-\dfrac{l}{L}\right)\sigma^2_{(S)(C)(L)}$ $+\left(1-\dfrac{r}{R}\right)\left(1-\dfrac{c}{C}\right)\sigma^2_{(S)(C)(RL)}$
$r \times c$ within sources $X_{(S)(RC)}$	$s(r-1)(c-1)$	$l\,\Sigma_{(S)(RC)} + \Sigma_{(S)(C)(RL)}$ $= l\,\sigma^2_{(S)(RC)} + \left(1-\dfrac{l}{L}\right)\sigma^2_{(S)(C)(RL)}$
strips within columns within sources $X_{(S)(C)(L)}$	$cs(l-1)$	$r\,\Sigma_{(S)(C)(L)} + \Sigma_{(S)(C)(RL)}$ $= r\sigma^2_{(S)(C)(L)} + \left(1-\dfrac{r}{R}\right)\sigma^2_{(S)(C)(RL)}$
(rows $\times$ strips) within columns within sources $X_{(S)(C)(RL)}$	$sc(r-1)(l-1)$	$\Sigma_{(S)(C)(RL)} = \sigma^2_{(S)(C)(RL)}$

145

TABLE 2. ANALYSIS OF VARIANCE FOR THE STRUCTURE $(S{:}Q)(P)$ AND $(SP{:}R)$

leading mean	d.f.	e.m.s.

$X_{(\theta)}$ 1

$$pqrs\ \Sigma_{(\theta)} + pqr\ \Sigma_{(S)} + qrs\ \Sigma_{(P)} + qr\ \Sigma_{(SP)} + pr\ \Sigma_{(S)(Q)}$$
$$+ r\Sigma_{(S)(PQ)} + q\ \Sigma_{(SP)(R)} + \Sigma_{(SP)(QR)}$$

$$= pqrs\ \sigma^2_{(\theta)} + pqr\left(1 - \frac{s}{i}\right)\sigma^2_{(S)} + qrs\left(1 - \frac{p}{P}\right)q^2_{(P)}$$

$$+ qr\left(1 - \frac{s}{S}\right)\left(1 - \frac{p}{P}\right)\sigma^2_{(SP)} + pr\left(1 - \frac{q}{Q}\right)\sigma^2_{(S)(Q)} + r\left(1 - \frac{p}{P}\right)\left(1 - \frac{q}{Q}\right)\sigma^2_{(S)(PQ)}$$

$$+ q\left(1 - \frac{r}{R}\right)\sigma^2_{(SP)(R)} + \left(1 - \frac{q}{Q}\right)\left(1 - \frac{r}{R}\right)\sigma^2_{(SP)(QR)}$$

$X_{(S)}$ $s-1$

$$pqr\ \Sigma_{(S)} + qr\ \Sigma_{(SP)} + pr\ \Sigma_{(S)(Q)} + r\ \Sigma_{(S)(PQ)} + q\ \Sigma_{(SP)(R)} + \Sigma_{(SP)(QR)}$$

$$= pqr\ \sigma^2_{(S)} + qr\left(1 - \frac{p}{P}\right)\sigma^2_{(SP)} + pr\left(1 - \frac{q}{Q}\right)\sigma^2_{(S)(Q)}$$

$$+ r\left(1 - \frac{p}{P}\right)\left(1 - \frac{q}{Q}\right)\sigma^2_{(S)(PQ)} + q\left(1 - \frac{r}{R}\right)\sigma^2_{(SP)(R)} + \left(1 - \frac{q}{Q}\right)\left(1 - \frac{r}{R}\right)\sigma^2_{(SP)(QR)}$$

$X_{(P)}$ $p-1$

$$qrs\ \Sigma_{(P)} + qr\ \Sigma_{(SP)} + r\ \Sigma_{(S)(PQ)} + q\ \Sigma_{(SP)(R)} + \Sigma_{(SP)(QR)}$$

$$= qrs\ \sigma^2_{(P)} + qr\left(1 - \frac{s}{S}\right)\sigma^2_{(SP)} + r\left(1 - \frac{q}{Q}\right)\sigma^2_{(S)(PQ)}$$

$$+ q\left(1 - \frac{r}{R}\right)\sigma^2_{(SP)(R)} + \left(1 - \frac{q}{Q}\right)\left(1 - \frac{r}{R}\right)\sigma^2_{(SP)(QR)}$$

$X_{(SP)}$ $(s-1)(p-1)'$

$$qr\ \Sigma_{(SP)} + r\ \Sigma_{(S)(PQ)} + q\ \Sigma_{(SP)(R)} + \Sigma_{(SP)(QR)}$$

$$= qr\ \sigma^2_{(SP)} + r\left(1 - \frac{q}{Q}\right)\sigma^2_{(S)(PQ)} + q\left(1 - \frac{r}{R}\right)\sigma^2_{(SP)(R)}$$

$$+ \left(1 - \frac{q}{Q}\right)\left(1 - \frac{r}{R}\right)_{(SP)(QR)}$$

$X_{(S)(Q)}$ $s(q-1)$

$$pr\ \Sigma_{(S)(Q)} + r\ \Sigma_{(S)(PQ)} + \Sigma_{(SP)(QR)} = pr\ \sigma^2_{(S)(Q)} + r\left(1 - \frac{p}{P}\right)\sigma^2_{(S)(PQ)}$$
$$+ \left(1 - \frac{r}{R}\right)\sigma^2_{(SP)(QR)}$$

$X_{(S)(PQ)}$ $s(q-1)(p-1)$

$$r\ \Sigma_{(S)(PQ)} + \Sigma_{(SP)(QR)} = r\ \sigma^2_{(S)(PQ)} + \left(1 - \frac{r}{R}\right)\sigma^2_{(SP)(QR)}$$

$X_{(SP)(R)}$ $sp(r-1)$

$$q\ \Sigma_{(SP)(R)} + \Sigma_{(SP)(QR)} = q\ \sigma^2_{(SP)(R)} + \left(1 - \frac{q}{Q}\right)\sigma^2_{(SP)(QR)}$$

$X_{(SP)(QR)}$ $sp(q-1(r-1)$

$$\Sigma_{(SP)(QR)} = \sigma^2_{(SP)(QR)}$$

TABLE 3.* ANALYSIS OF VARIANCE FOR A RANDOMIZED BLOCK DESIGN IN WHICH
q BLOCKS RECEIVE THE SAME ERROR FOR EACH TREATMENT LEVEL AND
w SUCH GROUPS OF BLOCKS ARE CHOSEN ALTOGETHER

Population structure and experimental procedure $(B : P)(G : g)$; sample structure

$(\ (i^*:m^*)\ (k^*):f)$; ambivalence relations $i^* \to B, g;\ m^* \to B;\ k^* \to G, P; f \to P$

source of variation and leading mean	e.m.s.
between groups of blocks x_{i^*}	$rg\ \Sigma_{(B)} + r\ \Sigma_{(BG)} + r\ \Sigma_{(G)(Bg)} + rq\ \Sigma_{(G)(g)} + \Sigma'$

$$= rg\ \sigma^2_{(B)} + r\left(1-\frac{g}{G}\right)\sigma^2_{(BG)} + \left(1-\frac{rg}{P}\right)\sigma^2_{(B)(P)}$$

$$+ \left(\left(1-\frac{1}{G}\right)-\frac{r}{P}\left(1-\frac{g}{G}\right)\right)\sigma^2_{(B)(PG)}$$

$$+ r\left(1-\frac{q}{B}\right)\sigma^2_{(G)(Bg)} + rq\ \sigma^2_{(G)(g)} + \left(1-\frac{r}{P}\right)\sigma^2_{(BG)(Pg)}$$

within groups of blocks $x_{i^*(m^*)}$ $\qquad rg\Sigma_{(B)} + r\ \Sigma_{(BG)} + r\ \Sigma_{(G)(Bg)} + \Sigma'$

$$= rg\ \sigma^2_{(B)} + r\left(1-\frac{g}{G}\right)\sigma^2_{(BG)} + \left(1-\frac{rg}{P}\right)\sigma^2_{(B)(P)}$$

$$+ \left(\left(1-\frac{1}{G}\right)-\frac{r}{P}\left(1-\frac{g}{G}\right)\right)\sigma^2_{(B)(GP)} + r\ \sigma^2_{(G)(Bg)}$$

$$+ \left(1-\frac{r}{P}\right)\sigma^2_{(BG)(Pg)}$$

between treatments x_{k^*} $\qquad rb\ \Sigma_{(G)} + r\ \Sigma_{(BG)} + r\ \Sigma_{(G)(Bg)} + rq\ \Sigma_{(G)(g)} + \Sigma'$

$$= rb\ \sigma^2_{(G)} + r\left(1-\frac{b}{B}\right)\sigma^2_{(BG)} + \sigma^2_{(B)(P)}$$

$$+ \left(\left(1-\frac{1}{G}\right)-\frac{r}{P}\right)\sigma^2_{(B)(PG)} + r\left(1-\frac{q}{B}\right)\sigma^2_{(G)(Bg)} + rq\ \sigma^2_{(G)(g)} + \left(1-\frac{r}{P}\right)\sigma^2_{(BG)(Pg)}$$

treatments $\times$ groups of blocks $x_{i^*k^*}$ $\qquad r\ \Sigma_{(BG)} + r\ \Sigma_{(G)(Bg)} + rq\ \Sigma_{(G)(g)} + \Sigma'$

$$= r\ \sigma^2_{(BG)} + \sigma^2_{(B)(P)} + \left(\left(1-\frac{1}{G}\right)-\frac{r}{P}\right)\sigma^2_{(B)(PG)}$$

$$+ r\left(1-\frac{q}{B}\right)\sigma^2_{(G)(Bg)} + rq\ \sigma^2_{(G)(g)} + \left(1-\frac{r}{P}\right)\sigma^2_{(BG)(Pg)}$$

treatments $\times$ blocks within groups $x_{i^*(m^*k^*)}$ $\qquad r\ \Sigma_{(BG)} + r\ \Sigma_{(G)(BG)} + \Sigma'$

$$= r\ \sigma^2_{(BG)} + \sigma^2_{(B)(P)} + \left(\left(1-\frac{1}{G}\right)-\frac{r}{P}\right)\sigma^2_{(B)(PG)} + r\ \sigma^2_{(G)(Bg)}$$

$$+ \left(1-\frac{r}{P}\right)\sigma^2_{(BG)(Pg)}$$

error $x_{i^*m^*k^*(f)}$ $\qquad \Sigma'$

$$= \sigma^2_{(B)(P)} - \left(1-\frac{1}{G}\right)\sigma^2_{(B)(PG)} + \sigma^2_{(BG)(Pg)}$$

* To shorten the size of the table the number of potential sub levels, M, within a treatment level, is taken to be infinite in all the σ^2 expressions for e.m.s.'s of the present table.

REFERENCES

ANSCOMBE, F. L. (1948) : The validity of comparative experiments. *J. Roy. Stat. Soc. Series A.*, 111, 181–200.

BENNETT, C. A. and FRANKLIN, N. L. (1954) : *Statistical Analysis in Chemistry and in the Chemical Industry*, John Wiley and Sons, Inc., New York.

BERKSON, J. (1950) : Are there two regressions? *J. Amer. Stat. Ass.*, 45, 164–180.

CORNFIELD, J. (1944) : On samples from finite populations. *J. Amer. Stat. Ass.*, 39, 236–239.

CORNFIELD, J. and TUKEY, J. W. (1956) : Average values of mean squares in factorials. *Ann. Math. Stat.*, 27, 907–949.

Cox, D. R. (1956) : A note on weighted randomization. *Ann. Math. Stat.*, 27, 1144–1151.

———— (1958) : The interpretation of the effects of non-additivity in the latin square. *Biometrika*, 45, 69–73.

EDEN, T. and YATES, F. (1933) : On the validity of Fisher's z-test when applied to an actual example of non-normal data. *J. Agr. Sci.*, 23, 6–16.

FISHER, R. A. (1918) : The correlation between relatives on the supposition of Mendelian inheritance. *Trans. Roy. Soc. Edin.*, 52, 339–433.

———— (1926) : The arrangement of field experiments. *J. Min. Agr. Eng.*, 33, 503–513.

———— (1930) : Moments and product moments of sampling distributions. *Proc. London Math. Soc.*, 30, 199–238.

———— (1935a) : *The Design of Experiments*, Oliver and Boyd, London.

KEMPTHORNE, O. (1952) : *The Design and Analysis of Experiments*, John Wiley and Sons, Inc., New York.

———— (1955) : The randomization theory of experimental inferences. *J. Amer. Stat. Ass.*, 50, 946–967.

KEMPTHORNE, O. and BARCLAY, W. D. (1953) : The partition of error in randomized blocks. *J. Amer. Stat. Ass.*, 48, 610–614.

NEYMAN, J., IWASZKIEWICZ, K. and ST. KOLODZIEYCZYK (1935) : Statistical problems in agricultural experimentation. *J. Roy. Stat. Soc. Suppl.*, 2, 107–154.

PITMAN, E. J. G. (1937) : Significance test which may be applied to samples from any population. III. The analysis of variance test. *Biometrika*, 29, 322–335.

WILK, M. B. (1955a) : Linear models and randomized experiments. *Unpublished Ph.D. Thesis.* Iowa State College Library, Ames, Iowa.

———— (1955b) : The randomization analysis of a generalized randomized block design. *Biometrika*, 42, 70–79.

WILK, M. B. and KEMPTHORNE, O. (1955) : Fixed, mixed, and random models. *J. Amer. Stat. Ass.*, 50, 1144–1167.

———— (1956a) : Some aspects of the analysis of factorial experiments in a completely randomized design. *Ann. Math. Stat.*, 27, 950–985.

———— (1956b) : Analysis of variance; preliminary tests, pooling, and linear models. WADC technical report, 55–244.

———— (1957) : Non-additivities in a latin square design. *J. Amer. Stat. Ass.*, 52, 218–236.

ZYSKIND, G. (1958) : Error structures in experimental designs. *Unpublished Ph.D. Thesis.* Iowa State College Library, Ames, Iowa.

ZYSKIND, G. and KEMPTHORNE, O. (1960) : Treatment errors in comparative experiments. WADC technical report, 59–19.

Paper received : July, 1960.

USE OF DISCRIMINANT AND ALLIED FUNCTIONS IN MULTIVARIATE ANALYSIS

By C. RADHAKRISHNA RAO

Indian Statistical Institute

SUMMARY. The use of discriminant function is justified only when it is known that the individual to be classified belongs to a certain subset of alternative groups. In this paper statistical methods are developed for examining whether a given individual can be supposed to have arisen from one of the populations in the subset and then using the discriminant function.

The use of certain other linear functions in problems of classification of groups, prediction, etc. has also been discussed.

1. INTRODUCTION

That the likelihood ratio is the best criterion for discrimination between two alternative hypotheses is justified by the fact that it is a sufficient statistic when only the two alternatives are considered (Smith, 1947). But, in practice, when it is not certain that an observed specimen belongs to one or the other of two given groups, there is need to consider the possibility of its belonging to some unknown group. The likelihood ratio which is equivalent to the discriminant function may not be sufficient when other alternatives are considered. Elsewhere (Rao, 1960), I have considered an example which gave rise to some controversy because such a possibility was ignored and a discriminant function was used to decide between two given alternatives (Bronowski and Long, 1951; Yates and Healy, 1951).

It is, therefore, important first to examine whether the discriminant function constructed on the basis of two given alternatives is *adequate* for drawing inference on an observed specimen. Only when this is confirmed by the observations on the specimen can we actually use the discriminant function for coming to a decision. The paper is devoted to development of suitable tests for this purpose.

Concepts of size and shape and their measurement based on multiple variables are also considered. These functions have been found useful in the study of inter-relationships and in the problem of discrimination between groups.

2. SUFFICIENCY OF THE DISCRIMINANT FUNCTION OVER A WIDER SET OF ALTERNATIVES

Let us consider two multivariate normal distributions with parameters (μ_1, Λ) and (μ_2, Λ) where μ_1, μ_2 are vectors of mean values and Λ, the common dispersion matrix. The log likelihood ratio, which is the linear discriminant function based on these two alternatives, is proportional to

$$\delta' \Lambda^{-1} x, \qquad \qquad \ldots \ (2.1)$$

where $\delta = \mu_1 - \mu_2$, Λ^{-1}, the reciprocal of Λ and x, the vector of observations. Lemma 1, which may be considered as a generalization of Smith's result, shows that the discriminant function (2.1) is not only sufficient for the alternatives from which it is derived but also for a wider set.

Lemma 1 : *The discriminant function $\delta' \Lambda^{-1} x$ is sufficient for the set of alternatives $(a_1 \mu_1 + a_2 \mu_2, \Lambda)$, where $a_1 + a_2 = 1$, i.e., the probability distribution of x given $\delta' \Lambda^{-1} x$ is the same for all expected values of x lying on the line joining μ_1 and μ_2.*

To prove the lemma it is enough to show that the ratio

$$P(x \mid a_1\mu_1 + a_2\mu_2) \div P(\delta' \Lambda^{-1} x \mid a_1\mu_1 + a_2\mu_2) \qquad \ldots \quad (2.2)$$

of normal densities is independent of (a_1, a_2) provided $a_1 + a_2 = 1$. The logarithm of (2.2) is proportional to

$$(x - a_1\delta - \mu_2)' \Lambda^{-1}(x - a_1\delta - \mu_2) - \frac{[\delta' \Lambda^{-1}(x - a_1\delta - \mu_2)]^2}{\delta' \Lambda^{-1}\delta} \qquad \ldots \quad (2.3)$$

where a_2 has been replaced by $(1 - a_1)$. The coefficients of a_1^2 and a_1 in (2.3) are obviously zero establishing the required result.

Lemma 2 which gives an extension of the result of Lemma 1 to the case of k multivariate normal populations with parameters (μ_i, Λ), $i = 1, \ldots, k$, can be proved on the same lines. Let $L_1, \ldots, L_{k-1}$ be $(k-1)$ independent likelihood ratios or $(k-1)$ independent discriminant functions arising out of some $(k-1)$ pairs of populations.

Lemma 2 : $L_1, \ldots, L_{k-1}$ are sufficient for the parameter set $a_1\mu_1 + \cdots + a_k\mu_k$ where $a_1 + \cdots + a_k = 1$.

3. Test prior to the application of discriminant function

As observed earlier, in any practical problem it is of some importance to consider the possibility of an observed specimen belonging to a group other than any one of those specified. For this purpose, it appears more natural first to test the hypothesis that the discriminant function is sufficient for drawing an inference on the observed specimen. Only if there is no evidence against this hypothesis could we use the discriminant function for further study, otherwise the question of classification of the observed specimen as a member of one of the two given groups does not arise. Again, we do not use the discriminant function just to decide between the two given alternatives only but also allow yet another possibility of the observed specimen belonging to a third group with its mean values lying on the line joining the mean values of the two given groups.

We shall first consider the case where all the parameters are known and the result of Lemma 1 regarding the sufficiency of the discriminant function is strictly valid.

The probability density of the observation x for an arbitrary mean μ is

$$P(x \mid \mu) = \text{const. } \exp\{-\tfrac{1}{2}(x - \mu)'\Lambda^{-1}(x - \mu)\}. \qquad \ldots \quad (3.1)$$

The criterion for testing the hypothesis that $\mu = a_1\mu_1 + a_2\mu_2$ where μ_1 and μ_2 are specified, is provided by the likelihood ratio,

$$-2 \log_e \frac{\sup\limits_{a_1 + a_2 = 1} P(x \mid a_1\mu_1 + a_2\mu_2)}{\sup\limits_{\mu} P(x \mid \mu)} \qquad \ldots \quad (3.2)$$

$$= \inf_{a_1} (x - \mu_2 - a_1\delta)' \Lambda^{-1} (x - \mu_2 - a_1\delta)$$

$$= (x - \mu_2)' \Lambda^{-1} (x - \mu_2) - \frac{[\delta' \Lambda^{-1} (x - \mu_2)]^2}{\delta' \Lambda^{-1} \delta} \qquad \ldots \quad (3.3)$$

$$= (x - \mu_1)' \Lambda^{-1} (x - \mu_1) - \frac{[\delta' \Lambda^{-1} (x - \mu_1)]^2}{\delta' \Lambda^{-1}\delta} \quad \text{(by symmetry)} \qquad \ldots \quad (3.4)$$

The statistic (3.3) is the difference between

$$\chi^2 = (x - \mu_2)' \, \Lambda^{-1} \, (x - \mu_2) \qquad \ldots \ (3.5)$$

which is a χ^2 on p d.f. and provides a test of the hypothesis that the individual belongs to the second population when nothing is known about the alternatives, and

$$\chi_1^2 = \frac{[\delta' \, \Lambda^{-1} \, (x - \mu_2)]^2}{\delta' \, \Lambda^{-1} \delta} \qquad \ldots \ (3.6)$$

which is a χ^2 on 1 d.f. and is useful for examining whether the individual belongs to the second population given that the alternatives for mean values are points on the line joining μ_1 and μ_2. We may then write the required test criterion (3.3) as

$$\chi_2^2 = \chi^2 - \chi_1^2 \qquad \ldots \ (3.7)$$

which is a χ^2 on $(p-1)$ d.f. If χ_2^2 is significantly large, there is evidence of the observed specimen belonging to a third group. On the other hand, if the value of χ_2^2 is not large, there is some assurance that the use of the discriminant function will not be misleading. The closeness of the observed specimen to one or the other of two groups is then judged by the discriminant function or same as chi-squares

$$\frac{[\delta' \, \Lambda^{-1} \, (x - \mu_2)]^2}{\delta' \, \Lambda^{-1} \delta} \quad \text{and} \quad \frac{[\delta' \, \Lambda^{-1} \, (x - \mu_1)]^2}{\delta' \, \Lambda^{-1} \delta} \qquad \ldots \ (3.8)$$

each with 1 d.f. As observed earlier, both these chi-squares may be large in a particular situation, indicating another possibility of the observed specimen belonging to a third group.

For example, in the classification of the Highdown Skull (Rao, 1952, pp. 294-296), considering the Bronze Age population we find $\chi^2 = 12.694$ and $\chi_1^2 = 3.993$ giving a difference, $\chi_2^2 = 8.701$ on 5 d.f. Considering the alternative of Maiden Castle population, $\chi^2 = 9.091$ and $\chi_1^2 = .390$ giving the same difference, $\chi_2^2 = 8.701$ on 5 d.f. which is not significantly high. We may then use χ_1^2 values on 1 d.f. for arriving at a decision. The χ_1^2 for Bronze Age is significant at the 5% level and that for Maiden Castle is small, which suggests a classification of the Highdown Skull as a member of the latter population.

Bronowski and Long (1952) amended the procedure, suggested by them in the earlier paper (1951), of using the discriminant function without a preliminary test of its sufficiency, to admit the possibility of an observed specimen belonging to a third group. Their new approach, however, does not make use of the discriminant function at all.

4. Tests when the number of alternatives is more than two

As before we consider the case where the alternative distributions are completely specified, i.e., the parameters involved are all known. In view of the result of Lemma 2, it may be examined first, whether $(k-1)$ independent likelihood ratios or discriminant functions arising out of k specified distributions are sufficient or not. Or in other words, whether the observed specimen can be considered to have arisen

from a population whose mean is on the hyper plane determined by the means of the k specified populations. If, as before, x stands for the vector of observations, $\mu_1, ..., \mu_k$ for mean values and Λ, the common dispersion matrix, the test criterion is

$$\chi_2^2 = \min_{\Sigma a = 1} (x - \Sigma a_i \mu_i)' \Lambda^{-1} (x - \Sigma a_i \mu_i) \qquad ... \quad (4.1)$$

which is a χ^2 on $(p-k+1)$ d.f. A method of computing a statistic of the form (4.1) is discussed in an earlier paper (Rao, 1959). If the value of χ_2^2 in (4.1) is not significantly high we consider only $(k-1)$ discriminant functions instead of p variables. For each group we then have a χ^2 on $(k-1)$ d.f. to judge the affinities of the observed specimen with that particular group.

5. Tests when the common dispersion matrix is estimated

In case Λ is estimated by S, which has Wishart's distribution on n d.f. the statistic

$$\frac{n-p+k}{n(p-k+1)} \, \chi_2^2 \qquad ... \quad (5.1)$$

which is constant times (4.1) with Λ^{-1} replaced by S^{-1}, has F distribution on $(p-k+1)$ and $(n-p+k)$ d.f. In the case $k = 2$, the F statistic is

$$\frac{n-p+2}{n(p-1)} \, \chi_2^2 = \frac{n-p+2}{n(p-1)} \left\{ (x-\mu_2)' S^{-1} (x-\mu_2) - \frac{[\delta' S^{-1}(x-\mu_2)]^2}{\delta' S^{-1} \delta} \right\} \qquad ... \quad (5.2)$$

with $p-1$ and $n-p+2$ d.f. These results follow from a general distribution problem solved in an earlier paper (Rao, 1959).

There is the further problem of testing the hypothesis that the specimen belongs to a particular group, using the estimated discriminant function or functions in the case of more than two alternatives. For $k = 2$, the test criterion for the hypothesis that the specimen belongs to (say) the second group is

$$\frac{n-p+1}{1} \left\{ \frac{[\delta' S^{-1}(x-\mu_2)]^2}{\delta' S^{-1} \delta} \div (n+\chi_2^2) \right\} \qquad ... \quad (5.3)$$

which is F on 1 and $(n-p+1)$ d.f. A similar statistic with an F distribution on $(k-1)$ and $(n-k+3)$ d.f. can be constructed in the case of k alternatives.

No exact tests of the types (5.2) and (5.3) seem to be available when the mean values of the two alternative populations are themselves estimated. An approximate test of the type (5.2) is provided by the smallest root of the determinantal equation defined by Fisher (1939). The further problem of using an estimated discriminant function as in (5.3) is not satisfactorily solved.

6. Size and shape functions

Penrose (1947) defined certain linear functions of measurements as representing the size and shape of an organism. More general functions were constructed by the author (Rao, 1958) to examine differences in size and shape between groups of organisms. These concepts are further generalised in this section.

Let us consider a linear function

$$X = a_1 x_1 + \ldots + a_k x_k + a_{k+1} x_{k+1} + \ldots + a_{k+m} x_{k+m} \qquad \ldots \text{(6.1)}$$

of $(k+m)$ variables $x_1, \ldots, x_{k+m}$, with unit variance, i.e.,

$$\Sigma \, \Sigma \, a_i a_j \lambda_{ij} = 1 \qquad \ldots \text{(6.2)}$$

where λ_{ij} is the covariance between x_i and x_j.

We wish to choose the coefficients $a_1, \ldots, a_{k+m}$ in such a way that a given increase in X produces on the average maximum changes, in given directions and given ratios, in a subset of the variables (say) $x_1, \ldots, x_s$. If the regression of x_i on X is taken to be linear then the average change in x_i for a unit change in X is exactly the regression coefficient of x_i on X. Let the ratios of specified changes in $x_1, \ldots, x_s$ including signs be $\rho_1 : \rho_2 : \ldots : \rho_s$. Using a multiplier α, we write the equations

$$\left. \begin{array}{ll} \Sigma \, a_i \lambda_{ij} = \alpha \, \rho_j, & j = 1, \ldots, s \\[2mm] \Sigma \, a_i \lambda_{ij} = \alpha \, \gamma_j, & j = 1, \ldots, k-s \end{array} \right\} \qquad \ldots \text{(6.3)}$$

where $\gamma_1, \ldots, \gamma_{k-s}$ are arbitrary constants. The algebraic problem is that of maximising α with the restriction (6.2) on a_i, by suitably choosing γ_j. Let Λ^{-1} be partitioned

$$\Lambda^{-1} = (\Lambda_1 \vdots \Lambda_2)$$

such that Λ_1 is of order $(k \times s)$, Λ_2 of order $(k \times \overline{s-k})$ and $\boldsymbol{\rho}$ and $\boldsymbol{\gamma}$ are the column vectors of elements ρ_i and γ_j respectively. The solution for $\boldsymbol{a}$, the column vector of a_i is

$$\boldsymbol{a} = (\Lambda_1 \vdots \Lambda_2)\left(\frac{\boldsymbol{\rho}}{\boldsymbol{\gamma}}\right) = \alpha(\Lambda_1 \boldsymbol{\rho} + \Lambda_2 \boldsymbol{\gamma}). \qquad \ldots \text{(6.4)}$$

Using (6.2) $\qquad \alpha^2 = 1 \div (\Lambda_1 \boldsymbol{\rho} + \Lambda_2 \boldsymbol{\gamma})' \Lambda (\Lambda_1 \boldsymbol{\rho} + \Lambda_2 \boldsymbol{\gamma}). \qquad \ldots \text{(6.5)}$

α is a maximum when the denominator in (6.5) is a minimum, i.e., $\boldsymbol{\gamma}$ is chosen to satisfy

$$\Lambda_2' \Lambda \Lambda_2 \boldsymbol{\gamma} = \Lambda_2' \Lambda \Lambda_1 \boldsymbol{\rho}. \qquad \ldots \text{(6.6)}$$

Having chosen $\boldsymbol{\gamma}$, we determine α from (6.5) and then $\boldsymbol{a}$ from (6.4) to obtain the desired linear function.

An alternative way of specifying the linear function (6.1) is as follows. The equations ensuring a given ratio of the regression coefficients of x_i on X, $i = 1, \ldots, s$ are

$$\Sigma \, a_i \lambda_{ij} = \alpha \, \rho_j, \quad j = 1, \ldots, s. \qquad \ldots \text{(6.7)}$$

The residual variance of x_i given x is

$$\lambda_{ii} - \alpha^2 \rho_i^2 / \boldsymbol{a}' \Lambda \boldsymbol{a}. \qquad \ldots \text{(6.8)}$$

We may choose $\boldsymbol{a}$, α, subject to restrictions (6.7) to minimise the sum of the residual variance for the s variables $x_1, \ldots, x_s$

$$\sum_1^s \left(\lambda_{ii} - \alpha^2 \rho_i^2 / \boldsymbol{a}' \Delta \boldsymbol{a}\right)$$

or maximise $\alpha^2 / \boldsymbol{a}' \Delta \boldsymbol{a}$. The equations leading to the determination of $\boldsymbol{a}$ are same as (6.4) and (6.6), thus establishing the equivalence of the two problems.

If ρ_i are all positive we may call (6.1) a size function with respect to the characters $x_1, \ldots, x_s$, in the sense that an increase in the value of size results in an increase, on the average, in each of the characters $x_1, \ldots, x_s$. By choosing some of the ρ_i to be positive and others negative we obtain a function with the property that an increase

in its value increases the value of some characters and decreases the value of some others. Such a function may be called a shape function.

It may be noted that in the construction of a size or a shape function for $x_1, ..., x_s$ we have used other measurements $x_{s+1}, ..., x_k$ also, and further that the linear function (6.1) constructed by the method indicated is invariant for linear transformations of the extra variables $x_{s+1}, ..., x_k$.

The present approach admits the possibility of constructing a linear function of any set of anthropometric measurements to represent shape of the head and to discriminate between long and broad headed people, instead of using the cephalic index. Such a linear function was used in an earlier paper (Rao, 1956) to study differences in head shape of castes and tribes belonging to different States in India.

A closely related problem is that of determining a linear function X of k variables such that the sum of the residual variances of $x_1, ..., x_s$ given X is a minimum, without placing any restrictions on the regression coefficients. Since the sum of residual variances is

$$\sum_{i=1}^{s} \left\{ \lambda_{ii} - \frac{(\Sigma \lambda_{ij} a_j)^2}{a'\Lambda a} \right\}$$

the problem is same as that of maximising

$$\sum_{i=1}^{s} \frac{(\Sigma \lambda_{ij} a_j)^2}{a'\Lambda a} = \frac{a'Ba}{a'\Lambda a} \qquad \qquad ... \quad (6.9)$$

where $B = A'A$, A being the $(s \times k)$ matrix obtained by taking the first s rows of the matrix Λ. The vector a maximising the ratio (6.9) is then the latent vector corresponding to the maximum latent root of the determinantal equation

$$|B - \lambda \Lambda| = 0. \qquad \qquad ... \quad (6.10)$$

In fact, by considering the first, second, third, ... latent vectors we obtain a series of linear functions $X_1, X_2, X_3, ...$ such that $X_1, ..., X_i$ (for any i) is an optimum set of i linear functions which can best predict the subset of variables $x_1, ..., x_s$. The determination of such functions is likely to be useful in specifying the sizes of ready-made garments, shoes, etc.

REFERENCES

Bronowski, J. and Long, W. M. (1951): Statistical methods in anthropology. *Nature*, **168**, 794.

———— (1952): *Amer. J. Phys. Anthrop.* 10. n.s. 385.

Fisher, R. A. (1939): The sampling distribution of some statistics obtained from non-linear regression. *Ann. Eugen.*, **9**, 238.

Penrose, L. S. (1947): Some notes on discrimination. *Ann. Eugen.*, **13**, 228–237.

Rao, C. R. (1952): *Advanced Statistical Methods in Biometric Research*, John Wiley and Sons, New York.

———— (1958): Appendix to Majumdar, D. N. and Rao, C. R. (1958): Bengal Anthropometric Survey, 1945: A Statistical Study. *Sankhyā*, **19**, 201.

———— (1959): Some problems involving linear hypothesis in multivariate analysis. *Biometrika*, **46**, 49.

———— (1960): Multivariate analysis: An indispensable aid in applied research. *Sankhyā*, **23**, 317–338.

Smith, C. A. B. (1947): Some examples of discrimination. *Ann. Eugen.*, **18**, 272–282.

Yates, F and Healy, M. J. R. (1951): Statistical methods in anthropology. *Nature*, **168**, 1116.

EQUALLY CORRELATED RANDOM VARIABLES*

By SIMEON M. BERMAN
Columbia University

SUMMARY. A representation theorem for a second order stationary sequence with constant covariance is given, and is applied to finding the limiting distribution of the maximum term in a sequence of equally correlated Gaussian random variable.

1. INTRODUCTION

A sequence $\{X_n : n \geqslant 1\}$ of random variables on a probability space $(\Omega, \mathcal{A}, P)$ is said to be equally correlated if

$$EX_n = 0, \quad EX_n^2 = 1 \quad (n \geqslant 1) \qquad \dots (1.1)$$
$$EX_n X_m = \rho \quad (m \neq n) \qquad \dots (1.2)$$

where $0 \leqslant \rho < 1$. The number ρ is restricted to the indicated range for if $\rho = 1$

$$X_1 = X_2 = \dots = X_n = \dots$$

with probability 1; the case $\rho < 0$ does not arise since the determinant of the covariance matrix of the first n variables,

$$\begin{vmatrix} 1 & \rho & \dots & & \rho \\ \rho & 1 & & & \\ \cdot & & \cdot & & \\ \cdot & & & 1 & \rho \\ \cdot & & & & \\ \rho & \rho & & \rho & 1 \end{vmatrix}$$

must be non-negative, i.e.

$$[n\rho + (1-\rho)](1-\rho)^{n-1} \geqslant 0 ;$$

hence

$$\rho \geqslant -\frac{1}{n-1}$$

for every n, so that ρ is non-negative. If $\rho = 0$, the random variables $\{X_n\}$ are orthogonal.

If $\{X_n\}$ is a sequence of exchangeable random variables, i.e. a sequence whose finite dimensional distributions are invariant under permutations of the indices, and if the first two moments exist and satisfy (1.1), then (1.2) holds, so that the $\{X_n\}$ are equally correlated. Exchangeable random variables have a well-known representation as "conditionally independent" random variables (Loeve, 1955). In this paper a representation for equally correlated random variables will be derived and will be used to get an explicit representation of exchangeable Gaussian variables.

2. THE REPRESENTATION THEOREM

Theorem : *Let* $\{X_n : n \geqslant 1\}$ *be a sequence of random variables satisfying* (1.1). *Then there exist a random variable* Y *and a sequence of random variables* $\{U_n\}$ *such that* $\qquad \dots (2.1)$
$$X_n = U_n + Y \quad (n \geqslant 1), \qquad \dots (2.2)$$
where
$$EU_n = EY = 0 ; \quad EU_n Y = 0$$
$$EU_n^2 = 1-\rho; \quad EY^2 = \rho; \quad EU_n U_m = 0 \quad (n \neq m).$$
If $\{X_n\}$ *is Gaussian, then* $Y, U_1, U_2, \dots$ *are Gaussian and independent.*

Proof : Since $\{X_n : n \geqslant 1\}$ is a wide sense stationary process, it follows from the ergodic theorem that there exists a random variable Y with a finite second moment such that

$$\text{l. i. m.} \frac{1}{n} \sum_{j=1}^{n} X_j = Y. \qquad \dots (2.3)$$

The sequence $\{U_n\}$ is now defined as

$$U_n = X_n - Y \qquad (n \geqslant 1); \qquad \qquad \ldots \ (2.4)$$

U_n is well-defined since X_n and Y are finite with probability one. The properties in (2.2) are immediate consequences of (2.3) and (2.4) and of the fact that mean square convergence implies the convergence of the first two moments; (2.1) follows from (2.4).

If $\{X_n\}$ is Gaussian, it follows from the proof given above that $Y, U_1, U_2, \ldots$ are jointly Gaussian; their independence then follows from the orthogonality relations in (2.2).

<div align="right">q.e.d.</div>

This theorem provides us with the *de* Finetti representation of the probability measure on a sequence of exchangeable Gaussian variables with equal correlation $\rho \geqslant 0$. The measure is of the form

$$P(.) = \int_{-\infty}^{\infty} P_y(\cdot) d\Phi(y/\sqrt{\rho}) \qquad \qquad \ldots \ (2.5)$$

where for each y, $P_y(\cdot)$ is the probability measure on a sequence of independent Gaussian random variables with mean y and variance $1-\rho$ and where $\Phi(u)$ is the standard normal integral.

3. Application to the distribution of the maximum

It was indicated in the first section that the correlation ρ between any pair of n equally correlated random variables must be greater than $-(n-1)^{-1}$, so that n such random variables are part of an infinite sequence of equally correlated random variables only if $\rho \geqslant 0$. The distribution of the maximum in a set of n equally correlated Gaussian variables with $\rho \geqslant -(n-1)^{-1}$ was computed from the inversion formula for the multivariate characteristic function by Kudo, (1958). Since the representation (2.5) is valid only for infinite sequences it is not applicable to the general problem considered by Kudo (1958); however, if we restrict ρ to nonnegative values (2.5) may be used. Then,

$$\max(X_1, \ldots, X_n) = \max(U_1 + Y, \ldots U_n + Y) = \max(U_1, \ldots, U_n) + Y, \qquad \ldots \ (3.1)$$

so that the distribution of $\max(X_1, \ldots, X_n)$ is simply the convolution of the distribution of the maximum of n *independent* Gaussian variables with the distribution of Y.

As $n \to \infty$, the limiting distribution of $\max(X_1, \ldots, X_n)$ after suitable normalization is the distribution of Y, i.e. Gaussian distribution. In fact, it has been shown (Gnedenko, 1943) that if $\{U_n; n \geqslant 1\}$ are independent Gaussian variables with mean zero and variance $1-\rho$, then

$$p \lim_{n \to \infty} [\max(U_1, \ldots, U_n) - \sqrt{2(1-\rho)\log n}\,] = 0.$$

Therefore, it follows from (3.1) that the distribution of

$$\max(X_1, \ldots, X_n) - \sqrt{2(1-\rho)\log n}$$

converges to the distribution of Y. On the other hand, if $X_1, X_2, \ldots X_n, \ldots$ are *independent* Gaussian variables, then the limiting distribution function of (3.1) is

$$\exp(-e^{-x}). \qquad \text{(Cramér, 1946)}.$$

References

Cramér, H. (1946) : *Mathematical Methods of Statistics*, Princeton University Press, 374.

Gnedenko, B. V. (1943) : Sur la Distribution Limite du Terme Maximum D'une Serie Aleatoire. *Ann. Math.*, **44**, 323-353.

Kudo, A. (1958) : On the distribution of the maximum value of an equally correlated sample from a normal population. *Sankhyā*, **20**, 309-316.

Loeve, M. (1955) : *Probability Theory*, Van Nostrand, New York, 365.

Paper received : October, 1961.

STUDENTISATION OF TWO-STAGE SAMPLE MEANS FROM NORMAL POPULATIONS WITH UNKNOWN VARIANCES*

I. GENERAL THEORY AND APPLICATION TO THE CONFIDENCE ESTIMATION AND TESTING OF THE DIFFERENCE IN POPULATION MEANS

By HAROLD RUBEN**

Columbia University

SUMMARY. A two-stage sampling procedure is developed for the estimation of means, and functions of means, of unknown normal populations by means of confidence sets of predetermined dimensions, and, concomitantly, for testing hypotheses relating to these means with power independent of the unknown variances. The estimation and test procedures are shown to be conservative and of high efficiency relative to the corresponding fixed sample size procedures when the variances are known.

The general theory is then used to obtain confidence intervals of preassigned confidence coefficient and length for the difference in means of two unknown normal populations, as well as tests for the equality of the means with power independent of the variances. The efficiency of these procedures for the two-mean problem is discussed in detail, with special emphasis on the minimisation of the expected total amount of sampling. Illustrative numerical examples with tables and graphs are provided.

1. INTRODUCTION

In a previous paper (Ruben, 1961) Stein's results (Stein, 1945) relating to the studentisation of sample means from normal populations which are assumed to have equal though unknown variances by means of two-stage sampling were derived in a different manner and in somewhat more extended and generalised form. There it was stressed that one of the advantages of the alternative method of approach is that it lends itself naturally to generalisation in the direction of the studentisation of sample means from normal populations with totally unknown variances. We propose here to carry through this programme.

Specifically, this series of papers investigates the problem of obtaining confidence sets with predetermined dimensions and confidence coefficient for the means, or for certain functions of the means, of a finite number of unconnected and unknown normal populations. (See also Ruben, 1950.) Preliminary samples are drawn which enable one to form estimates of the unknown variances, and these estimates determine a rule when to stop sampling. The sequential estimation and the corresponding similar test procedures involve the choice of constants n_{0i}, k_i which determine, respectively, the sizes of the preliminary samples and (approximately) the average total number of items sampled from the i-th population. These constants are chosen so as to meet the specifications laid down for the nature of the confidence sets or the form of the power functions.

A more significant interpretation of the constants or scale factors $1/k_i$ is that they play the same role in the present scheme of sequential sampling as the estimated standard errors of the means play in sampling with a predetermined number of items. This implies that the same scheme of sequential sampling allows one, in effect, to studentise the whole block of sample means from normal populations. Furthermore, in sampling from unknown normal populations with a predetermined number of items, the statistician is, as it were, at the mercy of the sampled values of the estimated standard errors of the means from the separate populations. It follows that the geometrical properties of his confidence sets and the power of his tests, even if available, are necessarily subject to change fluctuation in the one case and dependent on the unknown variances on the other. On the other hand, in sequential sampling what corresponds to the estimated standard errors of the means, viz. the $1/k_i$, may be fixed in advance. It follows that the geometrical properties of the confidence regions and the power of the tests may likewise be fixed in advance before sampling commences, independently of the unknown variances. In particular, confidence intervals of predetermined length and confidence coefficient, as well as tests of predetermined power may be obtained for the difference in means of two normal populations with unknown variances. This is discussed in some detail in the present paper.

It is shown that if the variance of the i-th population exceeds n_{0i}/k_i^2, then the efficiencies of the given sequential estimation procedures, as compared to the corresponding estimation procedures based on a predetermined number of observations when the variances are *known*, are nearly equal to unity, even when the n_{0i} are fairly small, and converge quite rapidly to unity as $n_{0i} \to \infty$. The question of improving on the efficiencies when the variances are small is briefly touched upon in the concluding paper.

2. Description of the Sampling Scheme

We suppose that there are $r(\geqslant 1)$ unconnected normal populations with unknown means and unknown variances, μ_i, σ_i^2 $(i = 1, 2, ..., r)$. Preliminary independent random samples of predetermined size n_{0i} are drawn from the populations, the results of such sampling being a set of observed values x_{ij} $(i, = 1, 2, ..., r;$ $j = 1, 2, ..., n_{0i})$. Further observations $x_{ij}(i = 1, 2, ..., r;$ $j = n_{0i}+1, n_{0i}+2, ..., n_i)$ are then obtained, where

$$n_i = \max (\{k_i^2 s_{0i}^2\}, n_{0i}) \qquad (i = 1, 2, ..., r), \qquad \text{...} \quad (2.1)$$

and

$$s_{0i}^2 = (n_{0i}-1)^{-1}\left[\sum_{j=1}^{n_{0i}} x_{ij}^2 -(\sum_{j=1}^{n_{0i}} x_{ij})^2/n_{0i}\right] \qquad (i = 1, 2, ..., r), \qquad \text{...} \quad (2.2)$$

the k_i being predetermined real constants with $\{q\}$ denoting the smallest integer not less than q. (The s_{0i}^2 are the pilot sample variance estimates based on $n_{0i}-1$ degrees of freedom). We shall refer to the scheme or rule of sampling described by (2.1) and (2.2) as $S(n_{01}, ..., n_{0r}; k_1, ..., k_r.)$ It will be noted that S amounts to applying a

Stein two-stage sampling procedure (Stein, 1945) *to each population separately.* The probability distributions of the n_i are therefore given by (*cf.* Stein, 1945 ; Ruben, 1961)

$$p_{m_i}(\sigma_i) = \begin{cases} 0 & (m_i < n_{0i}), \\ F_{\nu_{0i}}(n_{0i}h_i) & (m_i = n_{0i}), \\ F_{\nu_{0i}}(m_i h_i) - F_{\nu_{0i}}((m_i-1)h_i) & (m_i > n_{0i}), \end{cases} \quad \dots \ (2.3)$$

where
$$p_{m_i}(\sigma_i) = \Pr\{n_i = m_i | \mu_i, \sigma_i\}, \qquad \dots \ (2.4)$$
$$h_i = \nu_{0i}/k_i^2 \sigma_i^2, \qquad \dots \ (2.5)$$
$$\nu_{0i} = n_{0i} - 1, \qquad \dots \ (2.6)$$

and $F_{\nu_{0i}}(.)$ denotes the distribution function of χ^2 with ν_{0i} degrees of freedom. Furthermore, upper and lower bounds for En_i may be set as follows,

$$n_{0i}F_{\nu_{0i}}(n_{0i}h_i) + (\nu_{0i}/h_i)[1 - F_{\nu_{0i}+2}(n_{0i}h_i)] < En_i < n_{0i}F_{\nu_{0i}}(n_{0i}h_i)$$
$$+ (\nu_{0i}/h_i)[1 - F_{\nu_{0i}+2}(n_{0i}h_i)] + [1 - F_{\nu_{0i}}(n_{0i}h_i)], \qquad \dots \ (2.7)$$

whence, in particular, $\qquad En \sim k_i^2 \sigma_i^2, \qquad \dots \ (2.8)$

the approximation being reasonably good when $\nu_{0i}/h_i = k_i^2 \sigma_i^2 > n_{0i}$. The expected total amount of sampling is, of course, $E n_\cdot = \overset{r}{\underset{1}{\Sigma}} En_i$.

3. THE EXACT AND LIMITING PROBABILITY DISTRIBUTIONS OF THE TWO-STAGE SAMPLE MEANS AND OF FUNCTIONS OF THE MEANS

Denote the two-stage sample means by $\bar{x}_i$,

$$\bar{x}_i = \overset{n_i}{\underset{j=1}{\Sigma}} x_{ij}/n_i,$$

and the vector $(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_r)$ by $\boldsymbol{x}$. Similarly, the vectors $(\mu_1, \mu_2, \dots, \mu_r)$ and $(\sigma_1, \sigma_2, \dots, \sigma_r)$ will be denoted, for convenience, by $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ respectively, and the vectors $(n_1, n_2, \dots, n_r)$, $(m_1, m_2, \dots, m_r)$ by $\boldsymbol{n}$ and $\boldsymbol{m}$. Finally, the diagonal $r \times r$ matrix with diagonal elements σ_i^2/m_i $(i = 1, 2, \dots, r)$ will be denoted by $\boldsymbol{V}_{\boldsymbol{m}}$ and correspondingly the $r \times r$ matrix with diagonal elements $\sigma_i/\sqrt{m_i}$ by $\boldsymbol{V}_{\boldsymbol{m}}^{\frac{1}{2}}$. On indicating

$$\Pr\{\bar{x} \epsilon R | \boldsymbol{\mu}, \boldsymbol{\sigma}\},$$

for a given (Borel-measurable) region R in Euclidean r-space, by $P(R|\boldsymbol{\mu}, \boldsymbol{\sigma})$, we have

$$P(R|\boldsymbol{\mu}, \boldsymbol{\sigma}) = \overset{\infty}{\underset{m_1=1}{\Sigma}} \dots \overset{\infty}{\underset{m_r=1}{\Sigma}} p_{m_1}(\sigma_1) \dots p_{m_r}(\sigma_r) \Pr\{\bar{x} \epsilon R | \boldsymbol{n} = \boldsymbol{m}; \boldsymbol{\mu}, \boldsymbol{\sigma}\}. \quad \dots \ (3.1)$$

8

Again, the distribution of $\bar{x}$ conditional on n (say, $n = m$) is multivariate normal with expectation vector μ and variance-covariance matrix V_m. This follows from the independence of the vectors $(s_{01}^2, ..., s_{0r}^2)$ and $\left(\sum_{j=1}^{n_{01}} x_{1j}, ..., \sum_{j=1}^{n_{0r}} x_{rj} \right)$.

(The fixing of n is equivalent, by (2.1) to a constraint on the pilot sample variances which are distributed independently of the pilot sample means.) (3.1) then becomes

$$P(R | \mu, \sigma) = \sum_{m_1=1}^{\infty} ... \sum_{m_r=1}^{\infty} p_{m_1}(\sigma_1) ... p_{m_r}(\sigma_r) Q_m (R ; \mu, V_m^{\frac{1}{2}}), \quad ... \quad (3.2)$$

where

$$Q_m (R ; \mu, V_m^{\frac{1}{2}}) = \Pr \{\bar{x} \epsilon R | n = m ; \mu, \sigma\}$$

$$= (2\pi)^{-\frac{1}{2}r} \int ... \int_{\mu + V_m^{\frac{1}{2}} \xi \epsilon R} e^{-\frac{1}{2}\xi'\xi} d\xi. \quad ... \quad (3.3)$$

(3.2) and (3.3) give the required *exact* probability distribution of $\bar{x}$.

We shall now obtain the limiting probability distribution of $\bar{x}$, i.e.

$$\lim_{\sigma \to \infty} P(R | \mu, \sigma),$$

where it is understood that the $\sigma_i \to \infty$ $(i = 1, 2, ..., r)$ arbitrarily and independently of each other. For this purpose, let u_i $(i = 1, 2, ..., r)$ denote r independent χ^2 variates with ν_{0i} degrees of freedom, and note that

$$\frac{\sigma_i}{\sqrt{m_i}} = \frac{\sqrt{\nu_{0i}}/k_i}{\sqrt{m_i h_i}}, \quad ... \quad (3.4)$$

where h_i and ν_{0i} have been defined in (2.5) and (2.6), respectively. Denote further the $r \times r$ diagonal matrix with diagonal elements $\sqrt{\nu_{0i}}/k_i \sqrt{u_i}$ by U. Then, analogously to the case of equal σ_i^2 (see Ruben, 1961), the right-hand member of (3.2) may be identified as a Darboux sum associated with the r-fold integral

$$\int_{E_r^+} Q^*(R; \mu, U) \, dG \quad ... \quad (3.5)$$

and a specific divisions of Euclidean r-space. In (3.5),

$$Q^*(R; \mu, U) \equiv Q^*(R; \mu, \sqrt{\nu_{01}}/k_1\sqrt{u_1}, ..., \sqrt{\nu_{0r}}/k_r\sqrt{u_r})$$

$$= (2\pi)^{-\frac{1}{2}r} \int ... \int_{\mu + U\xi \epsilon R} e^{-\frac{1}{2}\xi'\xi} d\xi, \quad ... \quad (3.6)$$

E_r^+ denotes the positive orthant of Euclidean r-space, and G is the probability distribution in r-space generated by r independent χ^2 variates with ν_{0i} degrees of freedom $(i = 1, 2, ..., r)$, so that (3.5) is equivalent to

$$\int_0^{\infty} ... \int_0^{\infty} Q^*(R; \mu, \sqrt{\nu_{01}}/k_1\sqrt{u_1}, ..., \sqrt{\nu_{0r}}/k_r\sqrt{u_r})$$

$$\left(\prod_{i=1}^{r} \frac{1}{2^{\frac{1}{2}\nu_{0i}} \Gamma\left(\frac{\nu_{0i}}{2}\right)} e^{-\frac{1}{2}u_i} u_i^{\frac{1}{2}\nu_{0i}-1} \right) du_1...du_r. \quad ... \quad (3.7)$$

The division of r-space referred to is obtained by setting off points on the u_i-axis, distant $n_{0i}h_i$, $(n_{0i}+1)h_i$, $(n_{0i}+2)h_i$, ... from the origin and constructing through these points $(r-1)$-flats which are orthogonal to the axis, for $i = 1, 2, ..., r$. This results in generating an infinite set of orthotopes such that the probability-mass contained in the individual orthotopes is $\prod_1^r p_{m_i}(\sigma_i)$. (3.5) (or its equivalent (3.7)) then follows from (3.2), on using (3.4), replacing $m_i h_i$ by $\dot{u}_i$ and letting $h_i \to 0$. We thus have

$$P^*(R|\boldsymbol{\mu}) \equiv \lim_{\sigma \to \infty} P(R|\boldsymbol{\mu}, \boldsymbol{\sigma} = \int_{E_r^+} Q^*(\dot{R}; \boldsymbol{\mu}, U,) \, dG. \qquad \dots \quad (3.8)$$

(3.8) gives the required *limiting* probability distribution of $\bar{x}$.

In order to interpret further the probability in (3.8) consider the quantity

$$\Pr\{(\mu_1+t_{\nu_{01}}/k_1, ..., \mu_r+t_{\nu_{0r}}/k_r) \epsilon R\}, \qquad \dots \quad (3.9)$$

where $t_{\nu_{0i}}$ $(i = 1, ..., r)$ are independent Student random variables with ν_{0i} degrees of freedom. Introduce $2r$ independent random auxiliary variables ξ_i, u_i, where the ξ_i are normal with zero means and unit variances and the u_i (as before) are χ^2 variates with ν_{0i} degrees of freedom. Then (3.9) is equivalent to

$$\Pr\left\{ \left(\mu_1 + \frac{\xi_1}{k_1\sqrt{\dfrac{u_1}{\nu_{01}}}}, ..., \mu_r + \frac{\xi_r}{k_r\sqrt{\dfrac{u_r}{\nu_{0r}}}} \right) \epsilon R \right\}$$

$$= \int_{E_r^+} \Pr\left\{ \left(\mu_1 + \frac{\xi_1}{k_1\sqrt{\dfrac{u_1}{\nu_{01}}}}, ..., \mu_r + \frac{\xi_r}{k_r\sqrt{\dfrac{u_r}{\nu_{0r}}}} \right) \epsilon R \,\Big|\, u_1, ..., u_r \right\} dG, \qquad \dots \quad (3.10)$$

which is equal to the right-hand member of (3.8). (Recall the definition of Q^* in (3.6).) Recapitulating, for reference,

$$\lim_{\sigma \to \infty} \Pr\{\bar{x} \epsilon R | \boldsymbol{\mu}, \boldsymbol{\sigma}\} = \Pr\{(\mu_1+t_{\nu_{01}}/k_1, ..., \mu_r+t_{\nu_{0r}}/k_r) \epsilon R\}. \qquad \dots \quad (3.11)$$

In particular, the limiting distribution function $(\sigma \to \infty)$ of $\bar{x} - \boldsymbol{\mu}$ is identical with the distribution function of the vector $(t_{\nu_{01}}/k_1, ..., t_{\nu_{0r}}/k_r)$.

Finally, we wish to show that any statistical inference (such as confidence estimation or test procedure) based on (3.8), i.e. (3.8) is used to compute the appropriate probabilities (e.g. confidence coefficients or risks of errors) as *approximations* to the true probabilities will be conservative in character, provided Q^* satisfies certain monotonicity properties with respect to $u_1, ..., u_r$, or, equivalently, Q_m satisfies certain monotonicity properties with respect to $\sigma_1, ..., \sigma_r$. Specifically, we show that

$$\left. \begin{aligned} &\inf_{\sigma} P(R|\boldsymbol{\mu}, \boldsymbol{\sigma}) = \lim_{\sigma \to \infty} P(R|\boldsymbol{\mu}, \boldsymbol{\sigma}), \quad Q_m \text{ strictly decreasing} \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{in } \sigma_i \ (i = 1, 2, ..., r), \\[6pt] \text{and} \quad &\sup_{\sigma} P(R|\boldsymbol{\mu}, \boldsymbol{\sigma}) = \lim_{\sigma \to \infty} P(R|\boldsymbol{\mu}, \boldsymbol{\sigma}), \quad Q_m \text{ strictly increasing} \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{in } \sigma_i \ (i = 1, 2, ..., r). \end{aligned} \right\} \qquad \dots \quad (3.12)$$

To prove (3.12) it is sufficient merely to verify that the Darboux sum referred to previously in equ. (3.2) is an *upper* sum, relative to the previous cellular division of r-space, provided Q_m, defined in (3.3), is strictly decreasing in each of the σ_i (or, equivalently, Q^*, defined in (3.6), is strictly increasing in each of the u_i), and correspondingly the Darboux sum is a lower sum provided Q_m is strictly increasing in each of the σ_i (or, equivalently, Q^* is strictly decreasing in each of the u_i). We may then show from elementary properties of upper and lower Darboux sums, and analogously to the case where $\boldsymbol{\sigma}$ is one-dimensional (*cf.* Ruben, 1961), that

$$
\begin{aligned}
P(R|\boldsymbol{\mu}, \sqrt{C_1}\sigma_1, ..., \sqrt{C_r}\sigma_r) &< P(R|\boldsymbol{\mu}, \sigma_1, ..., \sigma_r), \quad Q_m \text{ strictly decreasing} \\
&\qquad \text{in the } \sigma_i \ (i = 1, 2, ..., r), \\
&> P(R|\boldsymbol{\mu}, \sigma_1, ..., \sigma_r), \quad Q_m \text{ strictly increasing} \\
&\qquad \text{in the } \sigma_i \ (i = 1, 2, ..., r),
\end{aligned}
\right\} * \quad (3.13)
$$

for all positive integral $C_i > 1$. (The increase of σ_i^2 by the factor C_i, i.e. the decrease of h_i by the factor C_i, has the effect of superimposing new points of division on the original division of r-space.) The existence of a value of $\boldsymbol{\sigma}$ for which $P(R|\boldsymbol{\mu}, \boldsymbol{\sigma})$ is less than $P^*(R|\boldsymbol{\mu})(Q_m$ decreasing in the σ_i), or for which $P(R|\boldsymbol{\mu}, \boldsymbol{\sigma})$ exceeds $P^*(R|\boldsymbol{\mu})$ $(Q_m$ increasing in the σ_i), then leads to an immediate contradiction. The greatest lower bound with respect to $\boldsymbol{\sigma}$ of $P(R|\boldsymbol{\mu}, \boldsymbol{\sigma})$ in the one case, and the least upper bound in the other case are thus attained asymptotically. Equ. (3.12) is thereby proved.

4. Estimation and testing

Consider the problem of obtaining symmetrical confidence intervals of preassigned width $2a$ and confidence coefficient $1-\alpha$ for the difference in means, $\mu_1-\mu_2$, of two normal populations with unknown variances σ_1^2, σ_2^2, and similarly the problem of testing $\mu_1-\mu_2$ with power function independent of σ_1^2 and σ_2^2 by means of the two-stage sampling scheme $S(n_{01}, n_{02}; k_1, k_2)$ described in Section 2. The problem of testing or estimating $\mu_1-\mu_2$ using samples of *fixed* size is the much discussed Behrens-Fisher problem. (See e.g. Fisher, 1935, 1941, 1956.) Fisher's solution (first given in essence by Behrens (1929)) is to compound the fiducial distributions of the two separate population means. Thus, we may write

$$
\mu_i = \bar{x}_i + \frac{s_i}{\sqrt{n_i}} \, t_{n_i-1} (i = 1, 2), ... \qquad ... \quad (4.1)
$$

where $\bar{x}_1$, $\bar{x}_2$ and s_1^2, s_2^2 are the two sample means and (unbiassed) variances based on samples of fixed size n_1 and n_2, while t_{n_1-1} and t_{n_2-1} denote independent Student variables with n_1-1 and n_2-1 degrees of freedom. We then have

$$
\frac{(\mu_1-\mu_2)-(\bar{x}_1-\bar{x}_2)}{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^{\frac{1}{2}}} = \sin\theta \, t_{n_1-1} - \cos\theta \, t_{n_2-1}, \qquad ... \quad (4.2)
$$

---

*It appears likely on intuitive grounds that $P(R|\boldsymbol{\mu}, \boldsymbol{\sigma})$ is strictly decreasing (increasing) in the σ_i if Q_m is strictly decreasing (increasing) in the σ_i. However, this awaits a formal proof.

where
$$\tan \theta = \frac{s_1/\sqrt{n_1}}{s_2/\sqrt{n_2}}. \qquad \qquad \ldots \text{(4.3)}$$

For purposes of comparison with the subsequent two-stage sample solution it is important to note that in Fisher's reasoning $\tan \theta$, the ratio of the estimated standard errors of the means, is regarded as constant (as are also the estimated means) once a particular pair of samples have been drawn from the two populations. In a sense then the field of probability inferences of $\mu_1 - \mu_2$ relate to a subset in which the s_i^2 as well as the $\bar{x}_i$ are held constant. Fisher's procedure for testing the null hypothesis $H_0 : \mu_1 = \mu_2$ on significance level α against the set of alternative hypotheses $\mu_1 \neq \mu_2$ when the σ_i are unknown is to base oneself on (4.2) and to reject H_0 when

$$\frac{|\bar{x}_1 - \bar{x}_2|}{\left(\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right)^{\frac{1}{2}}} \geqslant d_{n_1-1,\ n_2-1\ ;\ \theta\ ;\ \alpha/2}, \qquad \ldots \text{(4.4)}$$

where $d_{n_1-1,\ n_2-1\ ;\ \theta}$ is the weighted difference of two independent Student variables, as represented by the right-hand member of (4.2), and $d_{n_1-1,\ n_2-1\ ;\ \theta\ ;\ \alpha/2}$ is the two-sided $100\ \alpha\%$ significance point of $d_{n_1-1,\ n_2-1\ ;\ \theta}$. Similarly, an interval for $\mu_1 - \mu_2$ with fiducial probability $1-\alpha$ is

$$\left(\bar{x}_1 - \bar{x}_2 - d_{n_1-1,\ n_2-1\ ;\ \theta\ ;\ \alpha/2}\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^{\frac{1}{2}},\ \bar{x}_1 - \bar{x}_2 + d_{n_1-1,\ n_2-1\ ;\ \theta\ ;\ \alpha/2}\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^{\frac{1}{2}}\right). \ldots \text{(4.5)}$$

It is well known, however, that the legitimacy of Fisher's solution has been seriously questioned mainly on the grounds that θ cannot reasonably be regarded as constant after sampling for purposes of inference, and the application of Fisher's rule (4.4) when applied repeatedly to two fixed normal populations under identical conditions of sampling would not therefore lead to a proportion α of samples for which H_0 is rejected, when it is true, *independently of the value* σ_1/σ_2. (See e.g. Bartlett, 1936, 1937, 1956.)

5. CONFIDENCE ESTIMATION

Reverting now to $S(n_{01}, n_{02}; k_1, k_2)$, in order to generate confidence intervals of the prescribed type we need to base ourselves on the quantity

$$\Pr\{|(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)| < a\}, \qquad \ldots \text{(5.1)}$$

the region R of Section 3 being defined in the $(\bar{x}_1\ \bar{x}_2)$ plane by

$$R : |(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)| < a. \qquad \ldots \text{(5.2)}$$

From (3.11), the limiting distribution as $\sigma_1, \sigma_2 \to \infty$ of $(\bar{x}_1 - \mu_1) - (\bar{x}_2 - \mu_2)$ is the distribution of $t_{\nu_{01}}/k_1 - t_{\nu_{02}}/k_2$.

Consequently,

$$\frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\left(\frac{1}{k_1^2} + \frac{1}{k_2^2}\right)^{\frac{1}{2}}} \sim d_{\nu_{01},\ \nu_{02};\ \theta}, \qquad \dots (5.3)$$

where

$$\tan \theta = \frac{1}{k_1} \bigg/ \frac{1}{k_2} \qquad \dots (5.4)$$

and the symbol $\sim$ indicates the equality in distributions of the two random variables on either side of it as $\sigma_1, \sigma_2 \to \infty$. We note that θ is here strictly fixed and not subject to sampling fluctuation (cf. (5.4) with (4.3)). The required confidence intervals $(\bar{x}_1 - \bar{x}_2 - a, \ \bar{x}_1 - \bar{x}_2 + a)$ for $\mu_1 - \mu_2$, of width $2a$ and with limiting confidence coefficient $1 - \alpha$ are obtained directly from (5.3), and the four design constants n_{01}, n_{02}, k_1, k_2 must be related to the two specification constants a and α by

$$\left(\frac{1}{k_1^2} + \frac{1}{k_2^2}\right)^{\frac{1}{2}} d_{\nu_{01},\ \nu_{02};\ \theta;\ \alpha/2} = a, \qquad \dots (5.5)$$

but are otherwise arbitrary. The problem of a rational choice of the n_{0i} and k_i subject to the constraint (5.5) is discussed below.

It is of some interest to observe that sampling from two normal populations with the σ_i^2 regarded as having fiducial, or conceptual, distributions (distributions which have a subjective existence and arise from a thought process in the statistician's mind) obtained from an inversion argument applied to the chi-square distribution leads to a result which is equivalent to that obtained by sampling from two *fixed* normal populations, provided that the field of probability inference be restricted to samples whose sizes are conditioned by the relation (2.1).

The above estimation procedure is *conservative* in the sense that the actual (unknown) confidence coefficient, which is a function of the σ_i is *greater* than $1 - \alpha$. This follows from the first half of equ. (3.12) on noting that the value Q_m corresponding to (5.2) is

$$2\Phi[a(\sigma_1^2/m_1 + \sigma_2^2/m_2)^{-\frac{1}{2}}] - 1 \qquad \dots (5.6)$$

(where $\Phi(.)$ denotes, as usual, the distribution function of a normal variate with zero mean and unit variance), and is therefore strictly decreasing in σ_1 and σ_2. The true confidence coefficient is

$$\sum_{m_1 = 1}^{\infty} \sum_{m_2 = 1}^{\infty} p_{m_1}(\sigma_1) p_{m_2}(\sigma_2) \{2\Phi[a(\sigma_1^2/m_1 + \sigma_2^2/m_2)^{-\frac{1}{2}}] - 1\}. \qquad \dots (5.7)$$

Consider now the *efficiency* of the estimation procedure. According to (5.5) three degrees of freedom remain for the choice of the n_{0i} and k_i $(i = 1, 2)$. Even when the n_{0i} have been chosen there still remains an infinity of estimation procedures of our

type corresponding to the infinity of possible pairs of values for k_1 and k_2 satisfying (5.5). (This is to be contrasted to the sequential estimation of the mean of a single normal population (Stein, 1945; Ruben, 1961), where there is only one estimation procedure of our type which meets the given specifications, once the pilot sample size has been chosen.) A rational choice of the design constants would appear to be that corresponding to a minimisation of the expected total number of items sampled,[1] i.e. of

$$E(n_1+n_2) = \sum_1^\infty m_1 p_{m_1}(\sigma_1) + \sum_1^\infty m_2 p_{m_2}(\sigma_2), \qquad \dots \ (5.8)$$

around the suspected 'true' values of σ_1 and σ_2, subject to the constraint (5.5),[2] but we shall here, however, only determine an approximate minimisation in a number of specific cases. Observe first that by (5.5) and (5.4),

$$ak_1 = d_{\nu_{01},\ \nu_{02};\ \theta;\ \alpha/2}\ \mathrm{cosec}\ \theta, \quad ak_2 = d_{\nu_{01},\ \nu_{02};\ \theta;\ \alpha/2}\ \sec\ \theta, \qquad \dots \ (5.9)$$

so that on using formula (2.8),

$$E(n_1+n_2) \sim k_1^2\sigma_1^2 + k_2^2\sigma_2^2 = \left(\frac{d_{\nu_{01},\ \nu_{02};\ \theta,\ \alpha/2}}{a} \right)^2 (\sigma_1^2\ \mathrm{cosec}^2\theta + \sigma_2^2\ \sec^2\theta), \quad \dots \ (5.10)$$

the approximation being good when $k_i^2\sigma_i^2 > n_{0i}$. Now θ is completely at our disposal, since in the present scheme of sampling it is not a random variable but a definite constant. Obviously, in any particular case we should like to choose θ so as to minimise the right-hand member of (5.10). Unfortunately, the minimising value of θ is clearly a function of $\delta = \sigma_2^2/\sigma_1^2$, and δ being, by supposition, unknown the minimising value can never be exactly determined. Nevertheless, it appears reasonable to minimise $E(n_1+n_2)$ around the *suspected* value of δ, say δ_0.

It may be objected that the process of choosing a particular value of θ in advance to minimise, as far as possible, the labour of experimentation or observation is illogical since if δ were known, e.g. if δ were known to be δ_0, we should certainly not use the present procedure at all but rather one similar to that described elsewhere (Stein, 1945 ; Ruben, 1961). Indeed, as will be shown presently, this latter procedure is considerably more economical in sampling when $\delta = \delta_0$. However, we do not assume that $\delta = \delta_0$ strictly but are rather making use of what knowledge we have that δ may be *approximately* equal to δ_0. Even if this be quite false the procedure is not, of course, thereby invalidated. It simply means that the specifications could have been met with less experimentation than was actually the case. But it would be unreasonable not to use every possible scrap of information one might have, no theoretical grounds or otherwise, about δ.

---

[1] We assume here that the cost function is proportional to the number of items sampled.
[2] One of my students, Mr. H. Stein, has recently been pursuing this line of work by numerical methods, and it is hoped that his results will eventually appear in print.

We illustrate numerically with two examples, using equs. (5.9) to compute ak_1 and ak_2 and then the approximate formula (5.10) to compute $E(n_1+n_2)$. (A somewhat more precise, analytical discussion follows below.)

TABLE 1. THE (APPROXIMATE) EXPECTED SAMPLE SIZE

when $\nu_{01} = 6$, $\nu_{02} = 12$, $\alpha = 0.05$, for varying values of δ

$\theta°$	0	15	30	45	60	75	90
$a^2k_1^2$	∞	71.809	20.052	10.583	7.480	6.296	5.988
$a^2k_2^2$	4.748	5.153	6.682	10.583	22.440	87.666	∞
$E(n_1+n_2)a^2/\sigma_1^2$ when $\delta = 0.01$	∞	71.814	20.059	10.689	7.740	7.173	∞
$E(n_1+n_2)a^2/\sigma_1^2$ when $\delta = 0.1$	∞	72.324	20.720	11.641	9.724	15.062	∞
$E(n_1+n_2)a^2/\sigma_1^2$ when $\delta = 1$	∞	76.962	26.734	21.166	29.920	93.962	∞
$E(n_1+n_2)a^2/\sigma_1^2$ when $\delta = 5$	∞	97.574	53.462	63.498	119.680	444.626	∞
$E(n_1+n_2)a^2/\sigma_1^2$ when $\delta = 10$	∞	123.339	86.872	116.413	231.880	882.956	∞
$E(n_1+n_2)a^2/\sigma_1^2$ when $\delta = 100$	∞	587.109	688.252	1068.883	2251.480	8772.896	∞

TABLE 2. THE (APPROXIMATE) EXPECTED SAMPLE SIZE

when $\nu_{01} = 6$, $\nu_{02} = 12$, $\alpha = 0.01$, for varying values of δ

$\theta°$	0	15	30	45	60	75	90
$a^2k_1^2$	∞	139.3	38.56	21.086	15.93	14.18	13.74
$a^2k_2^2$	9.333	10.00	12.85	21.086	47.75	197.6	∞
$E(n_1+n_2)a^2/\sigma_1^2$ when $\delta = 0.01$	∞	139.4	38.69	21.30	16.41	16.16	∞
$E(n_1+n_2)a^2/\sigma_1^2$ when $\delta = 0.1$	∞	140.3	39.85	23.20	20.70	33.95	∞
$E(n_1+n_2)a^2/\sigma_1^2$ when $\delta = 1$	∞	149.3	51.41	42.13	63.67	211.8	∞
$E(n_1+n_2)a^2/\sigma_1^2$ when $\delta = 5$	∞	189.3	102.8	126.5	254.7	1002	∞
$E(n_1+n_2)a^2/\sigma_1^2$ when $\delta = 10$	∞	239.3	167.1	232.0	439.4	1990	∞
$E(n_1+n_2)a^2/\sigma_1^2$ when $\delta = 100$	∞	1139	1324	2130	4791	19780	∞

Both Tables show that $a^2k_1^2$ decreases steadily in θ, while $a^2k_2^2$ increases steadily with θ. The italicised numbers indicate the minimised values of $E(n_1+n_2)a^2/\sigma_1^2$ for the corresponding values of δ, as far as Sukhatme's Tables (Sukhatme, 1938), reproduced by Fisher and Yates (1957), of the percentile points of $d_{\nu_{01},\ \nu_{02};\ \theta}$ permit as to judge without interpolation. Thus, in Table 1, for $\delta = 10$, $E(n_1+n_2)a^2/\sigma_1^2$ decreases from ∞ when $\theta = 0°$ to a minimum value not far removed from 86.872 in the region

of $\theta = 30°$, and then increases to ∞ when $\theta = 90°$. If therefore there is some reason to believe that δ is more likely to have a value of about 10 than any other value, one would choose $\theta = 30°$ (about), i.e. one would choose values k_1 and k_2 for which $a^2 k_1^2 = 20.052$, $a^2 k_2^2 = 6.682$ in order to minimise the cost and labour of experimentation (e.g. if $a = 1$, $k_1^2 = 20.052$, $k_2^2 = 6.682$). Again, for $\delta = 10$ in the second example, the choice of θ would be again about 30° to minimise $E(n_1 + n_2) a^2 / \sigma_1^2$, leading to $a^2 k_1^2 = 38.56$, $a^2 k_2^2 = 12.85$ (e.g. if $a = 1$, $k_1^2 = 38.56$, $k_2^2 = 12.85$). [Note that the entries in the body of Table 2 are about double the corresponding entries in Table 1, so that the expected total sample size is nearly doubled when $\alpha = 0.01$ as compared to the previous case when $\alpha = 0.05$.]

We now examine the efficiency of the estimation procedure somewhat more rigorously. For purposes of discussion, it will be convenient and fruitful to schematize the problem of the estimation of the means of two separate normal populations by categorizing three successive stages of ignorance, as follows :

(i) Both σ_1^2 and σ_2^2 are known. This is hardly likely to arise in practice since one is not likely to know the variability of the populations and yet not know their locations.

(ii) The *ratio* σ_1^2 / σ_2^2 is known, though the separate values of σ_1^2 and σ_2^2 are not known. (A particularly important case is when σ_1^2 / σ_2^2 is known to be 1.)

(iii) Both σ_1^2 and σ_2^2 are unknown.

It is assumed that in all three cases the specifications to be met are given by the quantities $2a$ and α.

For (i), the standardised normal variate

$$\xi = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)^{\frac{1}{2}}} \qquad \dots \ (5.11)$$

is the most appropriate one to use, in the sense that it will minimise sampling. Here, n_1 and n_2 are *fixed* integers (there being as need for sequential sampling to obtain confidence intervals for $\mu_1 - \mu_2$ of fixed length when the σ_i^2 are *known*). To meet the specifications, the values of n_1 and n_2 must be chosen so as to satisfy the equation

$$\frac{a}{\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)^{\frac{1}{2}}} = \xi_{\alpha/2}, \qquad \dots \ (5.12)$$

where $\xi_{\alpha/2}$ is the two-sided $100\alpha\%$ significance point of ξ. There is still one degree of latitude available for the choice of n_1 and n_2. We should like to minimise the total number of items sampled,

$$n = n_1 + n_2, \qquad \dots \ (5.13)$$

subject to (5.12). The minimising values of n_1 and n_2 (using undetermined multipliers and regarding the n_i as continuous variables) are

$$n_i^* = \left(\frac{\xi_{a/2}}{a} \right)^2 \sigma_i(\sigma_1 + \sigma_2) \quad (i = 1, 2), \qquad \dots \quad (5.14)$$

whence the minimal (fixed) total sample size is

$$n^* \equiv n_1^* + n_2^* = \left(\frac{\xi_{a/2}}{a} \right)^2 (\sigma_1 + \sigma_2)^2. \qquad \dots \quad (5.15)$$

For (ii), we use the sequentially conditioned variate

$$\frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\left(\frac{1}{\rho_1^2} + \frac{\delta}{\rho_2^2} \right)^{\frac{1}{2}}} \sim t_{\nu_{01} + \nu_{02}}, \qquad \dots \quad (5.16)$$

where $t_{\nu_{01} + \nu_{02}}$ is a Student variable with $\nu_{01} + \nu_{02} = n_{01} + n_{02} - 2$ degrees of freedom and sampling is carried out according to the scheme

$$n_i = \max \left(\{ \rho_i^2 s_0^2 \}, n_{0i} \right) \qquad (i = 1, 2), \qquad \dots \quad (5.17)$$

with

$$s_0^2 = (n_{01} + n_{02} - 2)^{-1} \left[\sum_{j=1}^{n_{01}} x_{1j}^2 - \left(\sum_{j=1}^{n_{01}} x_{1j} \right)^2 \Big/ n_{01} \right]$$

$$+ (n_{01} + n_{02} - 2)^{-1} \delta^{-1} \left[\sum_{j=1}^{n_{02}} x_{2j}^2 - \left(\sum_{j=1}^{n_{02}} x_{2j} \right)^2 \Big/ n_{02} \right], \qquad \dots \quad (5.18)$$

s_0^2 being an unbiassed estimate of σ_1^2 on $\nu_{01} + \nu_{02}$ degrees of freedom. It is easy to show, using the method of a previous paper (Ruben, 1961), that the left-hand member of (5.16) is in fact distributed in the limit, as $\sigma_1^2 \to \infty$, in Student's form with $\nu_{01} + \nu_{02}$ degrees of freedom.

To meet the specifications, the scaling factors $1/\rho_1$ and $1/\rho_2$ which determine (approximately) the average total number of items sampled, denoted henceforth by $E_2 n$, must satisfy the equation

$$\frac{a}{\left(\frac{1}{\rho_1^2} + \frac{\delta}{\rho_2^2} \right)^{\frac{1}{2}}} = t_{\nu_{01} + \nu_{02}; \, \alpha/2} \qquad \dots \quad (5.19)$$

and accordingly

$$E_2 n \sim (\rho_1^2 + \rho_2^2) \sigma_1^2, \qquad \dots \quad (5.20)$$

provided only that σ_1^2 is not excessively small. We should like to minimise $E_2 n$ by the proper choice of ρ_1 and ρ_2, subject to the constraint (5.19). The minimising values are

$$\rho_1^* = \frac{t_{\nu_{01} + \nu_{02}; \, \alpha/2}}{a} \left(1 + \frac{\sigma_2}{\sigma_1} \right)^{\frac{1}{2}}, \quad \rho_2^* = \frac{t_{\nu_{01} + \nu_{02}; \, \alpha/2}}{a} \left[\frac{\sigma_2}{\sigma_1} \left(1 + \frac{\sigma_2}{\sigma_1} \right) \right]^{\frac{1}{2}} \dots \quad (5.21)$$

and the approximate minimal average total sample size is therefore

$$E_2^* n = \left(\frac{t_{\nu_{01}+\nu_{02};\, \alpha/2}}{a} \right)^2 (\sigma_1+\sigma_2)^2. \qquad \ldots \quad (5.22)$$

As for case (iii), this has already been discussed at some length previously. Denoting the total average sample size by $E_3(n\,|\,\delta)$, it has been shown in formula (5.10) that

$$E_3(n\,|\,\delta) \sim \left(\frac{d_{\nu_{01},\, \nu_{02};\, \theta;\, \alpha/2}}{a} \right)^2 \sigma_1^2 \, (\operatorname{cosec}^2\theta + \delta \sec^2\theta), \qquad \ldots \quad (5.23)$$

provided that the σ_i^2 are not excessively small, *where θ is an arbitrary constant which is entirely at our disposal.* For any assumed (suspected) value of δ, say $\delta = \delta_0$, there exists a minimising value of θ, say $\theta \equiv \theta(\delta_0) \equiv \theta_0$, i.e. a value of θ which will minimise $E_3(n\,|\,\delta_0)$. (The values of θ_0 corresponding to various assumed values of δ have been found *numerically* for two particular cases in Tables 1 and 2, as far as Sukhatme's Tables permitted without interpolation.) Analytically, the θ_0 corresponding to $\delta = \delta_0$ is a root of the transcendental equation

$$\frac{\partial}{\partial \theta} \, \{d^2_{\nu_{01},\, \nu_{02};\, \theta;\, \alpha/2} \, (\operatorname{cosec}^2\theta + \delta_0 \sec^2\theta)\} = 0. \qquad \ldots \quad (5.24)$$

A fairly good approximation to θ_0 is obtained from (5.24) by neglecting the variation of $d^2_{\nu_{01},\, \nu_{02};\, \theta;\, \alpha/2}$ with θ in comparison with the variation of $\operatorname{cosec}^2\theta + \delta_0 \sec^2\theta$ with θ. This is evidently justified if n_{01}, n_{02} are not too small, since $d_{\nu_{01},\, \nu_{02};\, \theta;\, \alpha/2} \to \xi_{\alpha/2}$, for *all* θ, as ν_{01}, $\nu_{02} \to \infty$.

Approximately then,

$$\theta_0 = \operatorname{arc\,cot} \delta_0^{1/4}. \qquad \ldots \quad (5.25)$$

Observe further that the approximation (5.25) is particularly good if δ_0 is either very small or very large. This follows from the fact that when $\delta_0 = 0$, or when $\delta_0 = \infty$, the solution given by (5.25) is entirely accurate. For, on setting $\delta_0 = 0$ in (5.24), one obtains

$$\frac{\partial}{\partial \theta} \, d_{\nu_{01},\, \nu_{02};\, \theta;\, \alpha/2} - d_{\nu_{01},\, \nu_{02};\, \theta;\, \alpha/2} \cot \theta = 0, \qquad \ldots \quad (5.26)$$

which is satisfied by $\theta = \pi/2$ on noting that

$$\frac{\partial}{\partial \theta} \, d_{\nu_{01},\, \nu_{02};\, \theta;\, \alpha/2} \Big|_{\theta=\pi/2} = 0. \qquad \ldots \quad (5.27)$$

Similarly, on setting $\delta_0 = \infty$ in (5.24), one obtains

$$\frac{\partial}{\partial \theta} \, d_{\nu_{01},\, \nu_{02};\, \theta;\, \alpha/2} + d_{\nu_{01},\, \nu_{02};\, \theta;\, \alpha/2} \tan \theta = 0, \qquad \ldots \quad (5.28)$$

which is satisfied by $\theta = 0$, on noting that

$$\frac{\partial}{\partial \theta}\, d_{\nu_{01},\, \nu_{02};\, \theta;\, \alpha/2}\Big|_{\theta\,=\,0} = 0. \qquad \cdots \quad (5.29)$$

((5.29) and (5.27) follow from the fact that $d_{\nu_{01},\, \nu_{02},\, -\theta}$ and $d_{\nu_{01},\, \nu_{02};\, \theta}$ are identically distributed variates, as are also $d_{\nu_{01},\, \nu_{02};\, (\pi/2)-\theta}$ and $d_{\nu_{01},\, \nu_{02};\, (\pi/2)+\theta}$, from the definition of $d_{\nu_{01},\, \nu_{02};\, \theta}$ in the form $d_{\nu_{01},\, \nu_{02};\, \theta} = \sin\theta\, t_{\nu_{01}} - \cos\theta\, t_{\nu_{02}}$. Therefore, $d_{\nu_{01},\, \nu_{02};\, \theta;\, \alpha/2}$, considered as a function of θ, is symmetrical round $\theta = 0$ and $\theta = \pi/2$.)

Equation (5.25) enables the minimising values of k_1 and k_2, k_1^* and k_2^*, to be determined approximately, and thereby also the minimal average total sample size at $\delta = \delta_0$, henceforth denoted by $E_3^*(n\,|\,\delta_0)$.

In fact, from (5.9) and (5.25),

$$\left.\begin{aligned}
k_1^* &= \frac{d_{\nu_{01},\, \nu_{02};\, \text{arc cot }\delta_0^{1/4};\, \alpha/2}}{a}\; (1+\delta_0^{\frac{1}{2}})^{\frac{1}{2}}, \\[2ex]
k_2^* &= \frac{d_{\nu_{01},\, \nu_{02};\, \text{arc cot }\delta_0^{1/4};\, \alpha/2}}{a}\; \delta_0^{-1/4}\,(1+\delta_0^{\frac{1}{2}})^{\frac{1}{2}},
\end{aligned}\right\} \qquad \cdots \quad (5.30)$$

so that
$$E_3^*(n\,|\,\delta) = \left(\frac{d_{\nu_{01},\, \nu_{02};\, \text{arc cot }\delta_0^{1/4};\, \alpha/2}}{a}\right)^2 \sigma_1^2(1+\delta_0^{\frac{1}{2}})(1+\delta\delta_0^{-\frac{1}{2}}) \qquad \cdots \quad (5.31)$$

(where $\delta = \sigma_2^2/\sigma_1^2$), and, in particular,

$$E_3^*(n\,|\,\delta_0) = \left(\frac{d_{\nu_{01},\, \nu_{02};\, \text{arc cot }\delta_0^{1/4};\, \alpha/2}}{a}\right)^2 (\sigma_1+\sigma_2)^2. \qquad \cdots \quad (5.32)$$

It is intuitively obvious that as our state of ignorance progressively increases from Case (i) to Case (iii), so does the cost and labour of experimentation, i.e.

$$n^* < E_2^* n < E_3^*(n\,|\,\delta_0), \qquad \cdots \quad (5.33)$$

where δ_0 has been substituted for δ in the expressions for n^* and $E_2^* n$. The author conjectures that

(a) no double-sample procedure meeting the given specifications could improve on that given by Case (ii) in which δ is known ($\delta = \delta_0$), at any rate provided only that sufficiently small values of σ_1^2 are excluded from consideration, and

(b) no double-sample procedure meeting the given specifications could improve on that given for Case (iii) in which σ_1^2 and σ_2^2 are both unknown and the procedure described is carried through on the assumption that $\delta = \delta_0$, if δ is indeed equal to δ_0, at any rate provided only that sufficiently small values of σ_i^2 are excluded from consideration.

Below we present a short illustrative Table which gives the approximate average amount of sampling in all three cases for various δ_0. The particular pair of pilot sample sizes has been chosen as (7, 13) and the two values $\alpha = 0.05$ and $\alpha = 0.01$ are considered. The approximate efficiency of the procedure given for Case (ii) is

$$Eff_2^* = \frac{n^*}{E_2^* n} = \left(\frac{\xi_{\alpha/2}}{t_{\nu_{01}+\nu_{02};\,\alpha/2}} \right)^2, \qquad \ldots \ (5.34)$$

independently of δ_0, while that for Case (iii) when $\delta = \delta_0$ is

$$Eff_3^*(\delta_0) = \frac{n^*}{E_3^*(n\,|\,\delta_0)} = \left(\frac{\xi_{\alpha/2}}{d_{\nu_{01},\nu_{02};\,\mathrm{arc\,cot}\,\delta_0^{1/4};\,\alpha/2}} \right)^2, \qquad \ldots \ (5.35)$$

Note in passing that
$$\lim_{\delta_0 \to 0} Eff_3^*(\delta_0) = \left(\frac{\xi_{\alpha/2}}{t_{\nu_{01};\,\alpha/2}} \right)^2, \qquad \ldots \ (5.36)$$

and
$$\lim_{\delta_0 \to \infty} Eff_3^*(\delta_0) = \left(\frac{\xi_{\alpha/2}}{t_{\nu_{02};\,\alpha/2}} \right)^2. \qquad \ldots \ (5.37)$$

The efficiencies have been defined in relation to n^*, since the latter provides, as it were, an absolute yardstick. No procedure, sequential or otherwise, which meets the stated requirements could hope to improve, in the sense of reducing average cost of sampling, on the procedure given for Case (i), whatever be the value of δ_0.

TABLE 3. APPROXIMATE AVERAGE AMOUNT OF SAMPLING AND EFFICIENCIES FOR $\nu_{01} = 6$, $\nu_{02} = 12$, $\alpha = 0.05$, 0.01, and for varying δ_0

δ_0	$n^* a^2/\sigma_1^2$		$E_2^*(n) a^2/\sigma_1^2$		$E_3^*(n\,|\,\delta_0) a^2/\sigma_1^2$		%Eff_2^*%		%$Eff_3^*(\delta_0)$%	
.01	4.649	8.030	5.862	10.02	7.173	16.17	87	80	65	50
.1	6.653	11.49	8.390	14.35	9.724	21.70	87	80	68	53
1	15.36	26.54	19.38	33.13	21.17	42.18	87	80	73	63
5	40.23	69.49	50.73	86.73	53.46	102.8	87	80	75	68
10	66.53	114.9	83.90	143.5	86.87	167.1	87	80	77	69
100	464.9	803.0	586.2	1002	587.1	1139	87	80	79	70
∞	∞	∞	∞	∞	∞	∞	87	80	81	71

(The entries on the left-hand side of each column after the first correspond to $\alpha = 0.05$ and the entries on the right to $\alpha = 0.01$.)

We observe that in this particular case $Eff_3^*(\delta_0)$ is monotonic in δ_0, both when $\alpha = 0.05$ and $\alpha = 0.01$. From (5.35), we see that this is due to the fact that both $d_{6,12;\,\theta_0;\,0.025}$ and $d_{6,12;\,\theta_0;\,0.005}$ increase steadily with θ_0. However, this will not

always be the case. Fisher (1941) has pointed out that the manner of variation of the percentile points of the d-distribution may be considerably affected by the significance level (in our case, the confidence coefficient) on which one works. For example, $d_{12,8;\ \theta_0;\ 0.025}$ decreases steadily with θ_0 for $0 < \theta_0 < \pi/2$; on the other hand, $d_{12,8;\ \theta_0;0.005}$ is a single-humped function of θ_0 in the range $0 < \theta_0 < \pi/2$. Consequently $Eff_3^*(\delta_0)$ will not be a monotonic function of δ_0 when $n_{01} = 13$, $n_{02} = 9$ and $\alpha = 0.01$.

So far we have considered only the efficiency of the procedure when $\delta = \delta_0$. However, for a completely adequate characterisation of the procedure we need to know its efficiency for *all* possible δ. This will indicate how the efficiency falls off as δ departs from the suspected value δ_0 at which it clearly attains its maximum value. From (5.31) and (5.15), the efficiency as a function of δ is given approximately by

$$Eff_3^*(\delta) = \left(\frac{\xi_{\alpha/2}}{\delta_{\nu_{01},\ \nu_{02};\ \text{arc cot}\ \delta_0^{1/4};\ \alpha/2}} \right)^2 (1+\delta_0^{\frac{1}{2}})^{-1}(1+\delta\delta_0^{-\frac{1}{2}})^{-1}(1+\delta^{\frac{1}{2}})^2 \quad \dots \quad (5.38)$$

The efficiency curve is single-humped. It rises with infinite slope at $\delta = 0$ where the efficiency is

$$\left(\frac{\xi_{\alpha/2}}{d_{\nu_{01},\ \nu_{02};\ \text{arc cot}\ \delta_0^{\frac{1}{4}};\ \alpha/2}} \right)^2 (1+\delta_0^{\frac{1}{2}})^{-1} \quad \dots \quad (5.39)$$

to the maximum value $Eff_3^*(\delta_0)$, as given in (5.35), at $\delta = \delta_0$; it then falls asymptotically to the value

$$\left(\frac{\xi_{\alpha/2}}{d_{\nu_{01},\ \nu_{02};\text{arc cot}\ \delta_0^{1/4};\ \alpha/2}} \right)^2 \delta_0^{\frac{1}{2}}(1+\delta_0^{\frac{1}{2}})^{-1}. \quad \dots \quad (5.40)$$

It will be observed that the process is not asymptotically $(n_{01}, n_{02} \to \infty)$ efficient *uniformly* (i.e. for all δ), but only for the single value $\delta = \delta_0$. The reason for this state of affairs is essentially as follows. For Case (i),

$$\frac{n_2^*}{n_1^*} = \frac{\sigma_2}{\sigma_1} = \delta^{\frac{1}{2}}. \quad \dots \quad (5.41)$$

However, for Case (iii) our procedure gives, approximately, for the ratio of the two optimal average sample sizes,

$$\frac{E_3^* n_2}{E_3^* n_1} = \frac{k_2^2}{k_1^2} \frac{\sigma_2^2}{\sigma_1^2} = \delta \tan^2\theta_0 = \delta\delta_0^{-\frac{1}{2}}. \quad \dots \quad (5.42)$$

Thus, the left-hand members of (5.41) and (5.42) are equal if, and only if, $\delta = \delta_0$. There is a degree of latitude available in the fixed sampling size case (when the σ_i^2 are known) enabling one to allocate the sizes of the two samples in a certain optimal ratio which is lacking in the variable sampling size case (when the σ_i^2 are unknown). If,

however, we relate the efficiency of the procedure for Case (iii) not to the absolutely optimum procedure for Case (i), characterised by the choice of sample sizes $n_1 = n_1^*$, $n_2 = n_2^*$, but rather to the sub-sequence of repeated experiments for Case (i) defined by

$$\frac{n_2}{n_1} = \frac{E_3^* n_2^*}{E_3^* n_1^*} = \delta \delta_0^{\frac{1}{2}}, \qquad \qquad \dots \quad (5.43)$$

then the efficiency is now

$$\overline{Eff_3^*}(\delta) = \left(\frac{\xi \alpha/2}{d_{\nu_{01}, \nu_{02}; \text{ arc cot } \delta_0^{1/4}; \alpha/2}} \right)^2 = Eff_3^*(\delta_0), \qquad \dots \quad (5.44)$$

for *all* δ. This follows from (5.12) on substituting

$$\frac{n_2}{n_1} = \delta \delta_0^{-\frac{1}{2}}, \qquad \qquad \dots \quad (5.45)$$

for then $\qquad n \doteq n_1 + n_2 = \left(\frac{\xi \alpha/2}{a} \right)^2 \sigma_1^2 (1 + \delta_0^{\frac{1}{2}})(1 + \delta \delta_0^{-\frac{1}{2}}), \qquad \dots \quad (5.46)$

and comparing (5.46) with (5.31) the 'conditional' efficiency is verified as that given in (5.44), no matter what is the true value of δ.

We illustrate further by a particularly important example. Let $n_{01} = n_0 = n_{02}$, and suppose that there is external evidence for believing that δ may be not far removed from $1 (\delta_0 = 1)$. Now since $d_{\nu_0, \nu_0; (\pi/4) - \theta}$ and $d_{\nu_0, \nu_0; (\pi/4) + \theta}$ are identically distributed, $d_{\nu_0, \nu_0; \theta; \alpha/2}$, considered as a function of θ, is symmetrical round $\theta = \pi/4$, therefore

$$\left. \frac{\partial}{\partial \theta} d_{\nu_0, \nu_0; \theta; \alpha/2} \right|_{\theta = \pi/4} = 0,$$

where here $\nu_0 = n_0 - 1$.

Also, $\qquad \qquad \left. \frac{\partial}{\partial \theta} (\operatorname{cosec}^2 \theta + \sec^2 \theta) \right|_{\theta = \pi/4} = 0.$

Consequently, for $\delta_0 = 1$, (5.24) is satisfied by $\theta_0 = 45°$.* (Note that here $\theta_0 = 45°$, as given by (5.25), is the *exact* value of the root.) On referring to (5.30), $k_1^* = k_2^* = k^*$ (say), where

$$k^* = \sqrt{2} \left(\frac{d_{\nu_0, \nu_0; 45°; \alpha/2}}{a} \right). \qquad \dots \quad (5.47)$$

From (5.31) and (5.38), we have (again accurately in this case)

$$E_3^* (n|\delta) = 2 \sigma_1^2 \left(\frac{d_{\nu_0, \nu_0; 45°; \alpha/2}}{a} \right)^2 (1 + \delta), \qquad \dots \quad (5.48)$$

and $\qquad \qquad Eff_3^*(\delta) = \frac{1}{2} \left(\frac{\xi \alpha/2}{d_{\nu_0, \nu_0; 45°; \alpha/2}} \right)^2 (1 + \delta)^{-1} (1 + \delta^{\frac{1}{2}})^2. \qquad \dots \quad (5.49)$

---

*The distribution of $d_{\nu_0, \nu_0; \pi/4}$ has been found elsewhere as a series expansion (Ruben, 1960).

$Eff_3^*(\delta)$ is graphed in Figs. 1 and 2 for various values of n_0 and for $\alpha = 0.05$, $\alpha = 0.01$. In particular,

$$Eff_3^*(1) = \left(\frac{\xi_{\alpha/2}}{d_{\nu_0, \nu_0; 45°; \alpha/2}} \right)^2, \qquad \ldots \quad (5.50)$$

giving the maximum height of the efficiency curve. The procedure is asymptotically efficient $(n_0 \to \infty)$ if $\delta = 1$, but not otherwise. However, the procedure is fairly nearly asymptotically efficient in the range $0.1 < \delta < 10$: in this range the efficiency is asymptotically not less than about 79%.

In relation to the subsequence of experiments for Case (i), in which as from (5.43), $n_2/n_1 = \delta$, the efficiency is, by (5.44),

$$Eff_3^*(\delta) = \left(\frac{\zeta_{\alpha/2}}{d_{\nu_0, \nu_0; 45°; \alpha/2}} \right)^2 = Eff_3^*(1). \qquad \ldots \quad (5.51)$$

Again, for purposes of comparison with Case (ii), note that when $\delta = 1$ (on using (5.21), (5.17) and (5.18)),

$$\rho_1^* = \rho^* = \rho_2^*, \qquad \ldots \quad (5.52)$$

with

$$\rho^* = \sqrt{2} \, t_{2(n_0-1); \, \alpha/2} | a, \qquad \ldots \quad (5.53)$$

and

$$n_1 = \max (\{\rho^{*2} s_0^2\}, n_0) = n_2, \qquad \ldots \quad (5.54)$$

TABLE 4(a)

$n_0 - 1$	'conditional' efficiency $\overline{Eff_3^*(1)}\%$
6	65
8	73
12	82
24	91
∞	100

$\alpha = 0.05$

TABLE 4(b)

$n_0 - 1$	'conditional' efficiency $Eff_3^*(1)\%$
6	54
8	65
12	76
24	88
∞	100

$\alpha = 0.01$

with

$$s_0^2 = \frac{1}{2(n_0-1)} \left[\sum_{i=1}^{2} \sum_{j=1}^{n_0} x_{ij}^2 - \frac{1}{n_0} \left(\left(\sum_{j=1}^{n_0} x_{1j} \right)^2 + \left(\sum_{j=1}^{n_0} x_{2j} \right)^2 \right) \right]. \qquad \ldots \quad (5.55)$$

Furthermore,

$$Eff_2^* = \left(\frac{\xi_{\alpha/2}}{t_{2(n_0-1); \, \alpha/2}} \right)^2, \qquad \ldots \quad (5.56)$$

which is to be compared with (5.51). This procedure (for Case (ii)) is now a member of the class of procedures discussed previously (Stein, 1945; Ruben, 1961) where the sequential testing of the general linear hypothesis and the concomitant estimation problem are considered.*

---

*Reference should also be made to a paper by Chapman (1950) in which an analogous two-stage procedure is developed for the confidence estimation of the difference in means in a rather special case.

Fig. 1

Fig. 2
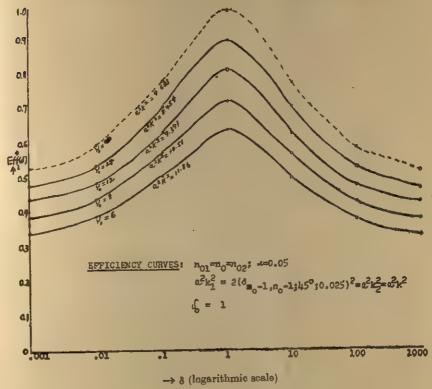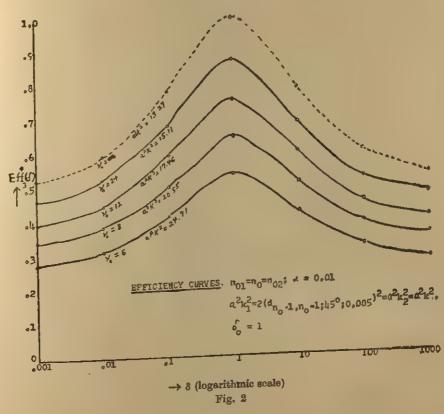
6. The testing of the difference in means of two unknown normal populations

For convenience, denote $\mu_1 - \mu_2$ by η. We wish to test η so that the power function is independent of the unknown σ_i^2. For this purpose, the two-stage sampling procedure $S(n_{01}, n_{02}; k_1, k_2)$ will be used.

Consider then the problem of testing the simple hypothesis $H_0 : \eta = \eta_0$ against the simple alternative $H_1 : \eta = \eta_1$, with $\eta_1 > \eta_0$. We assume that the risks of the two types of errors are specified in advance as α and β. The rejection criterion

$$R : \bar{x}_1 - \bar{x}_2 \geqslant C \qquad \qquad \dots \quad (6.1)$$

will be used for H_0, where the five constants n_{01}, n_{02}, k_1, k_2 and C have to be chosen so as to satisfy the imposed requirements. From (5.3), the limiting power function of the test is

$$1 - \psi_{\nu_{01}, \, \nu_{02}; \, \theta} \left[\frac{C - \eta}{\left(\frac{1}{k_1^2} + \frac{1}{k_2^2} \right)^{\frac{1}{2}}} \right] \qquad \dots \quad (6.2)$$

where $\psi_{\nu_{01}, \, \nu_{02}; \, \theta}(\cdot)$ is the distribution function of $d_{\nu_{01}, \, \nu_{02}; \, \theta}$. We therefore require

$$1 - \psi_{\nu_{01}, \, \nu_{02}; \, \theta} \left[\frac{C - \eta_0}{\left(\frac{1}{k_1^2} + \frac{1}{k} \right)^{\frac{1}{2}}} \right] = \alpha \qquad \dots \quad (6.3)$$

and

$$\psi_{\nu_{01}, \, \nu_{02}; \, \theta} \left[\frac{C - \eta_1}{\left(\frac{1}{k_1^2} + \frac{1}{k_2^2} \right)^{\frac{1}{2}}} \right] = \beta, \qquad \dots \quad (6.4)$$

whence

$$\frac{C - \eta_0}{\left(\frac{1}{k_1^2} + \frac{1}{k_2^2} \right)^{\frac{1}{2}}} = d_{\nu_{01}, \, \nu_{02}; \, \theta; \, \alpha} \qquad \dots \quad (6.5)$$

and

$$\frac{C - \eta_1}{\left(\frac{1}{k_1^2} + \frac{1}{k_2^2} \right)^{\frac{1}{2}}} = -d_{\nu_{01}, \, \nu_{02}; \, \theta; \, \beta}, \qquad \dots \quad (6.6)$$

or, equivalently (recall that $\tan \theta = k_2/k_1$),

$$(C - \eta_0) k_1 = d_{\nu_{01}, \, \nu_{02}; \, \theta; \, \alpha} \operatorname{cosec} \theta, \qquad \dots \quad (6.7)$$

and $\qquad (C - \eta_1) k_1 = -d_{\nu_{01}, \, \nu_{02}, \, \theta; \, \beta} \operatorname{cosec} \theta. \qquad \dots \quad (6.8)$

Solving (6.7) and (6.8) for C and k_1,

$$C = \frac{d_{\nu_{01}, \, \nu_{02}; \theta; \alpha} \, \eta_1 + d_{\nu_{01}, \, \nu_{02}; \theta; \beta} \, \eta_0}{d_{\nu_{01}, \, \nu_{02}; \theta; \alpha} + d_{\nu_{01}, \, \nu_{02}; \theta; \beta}}, \qquad \dots \quad (6.9)$$

$$k_1 = \frac{(d_{\nu_{01}, \, \nu_{02}; \, \theta; \, \alpha} + d_{\nu_{01}, \, \nu_{02}; \, \theta; \, \beta}) \operatorname{cosec} \theta}{\eta_1 - \eta_0} \qquad \dots \quad (6.10)$$

so that $\qquad k_2 = \dfrac{(d_{\nu_{01}, \, \nu_{02}; \, \theta; \, \alpha} + d_{\nu_{01}, \, \nu_{02}; \, \theta; \, \beta}) \sec \theta}{\eta_1 - \eta_0}. \qquad \dots \quad (6.11)$

For given $\alpha, \beta, \eta_0, \eta_1$, the constants n_{01}, n_{02}, k_1 and k_2 may be chosen arbitrarily, provided only that the latter satisfy equs. (6.10) and (6.11). For each such choice the

rejection criteria is provided by (6.1) and (6.9). If further the pilot sample sizes are selected in advance then there remains only to choose θ, and k_1, k_2 are automatically determined by (6.10) and (6.11) after θ has been chosen. For all such θ, the specified requirements for the risks of error will be met. In practice, one would, exactly as for the sequential estimation procedure, choose the value of θ which will minimise $En \sim k_1^2 \sigma_1^2 + k_2^2 \sigma_2^2$ when $\delta = \sigma_2^2/\sigma_1^2$ has some suspected value.

The power function in (6.2) is monotonic in $\mu_1 - \mu_2$. This implies that the present procedure may be used to test the more realistic composite hypothesis H_0': $\eta \leqslant 0$ against $\eta > 0$, when it is specified that the power $\leqslant \alpha$ for $\eta \leqslant$ some value η_0 of $\eta(\eta_0 < 0)$, and $\geqslant 1-\beta$ for $\eta \geqslant$ a value η_1 of $\eta(-\eta_0 = \eta_1 > 0)$, as well as to test the hypothesis H_0'': $\eta = 0$ against one-sided alternatives $\eta > 0$, when it is specified that the probability of incorrectly rejecting the null hypothesis shall be α and the probability of correctly rejecting shall be $\geqslant 1-\beta$ for $\eta \geqslant \eta_1$, η_1 being some suitably chosen (positive) value of η. For the first of these two cases (6.9), (6.10) and (6.11) reduce to

$$C = \left[\frac{d_{\nu_{01}, \nu_{02}; \theta; \alpha} - d_{\nu_{01}, \nu_{02}; \theta; \beta}}{d_{\nu_{01}, \nu_{02}, \theta, \alpha} + d_{\nu_{01}, \nu_{02}, \theta, \beta}} \right] n_1, \qquad \ldots \quad (6.12)$$

$$k_1 = (d_{\nu_{01}, \nu_{02}; \theta; \alpha} + d_{\nu_{01}, \nu_{02}; \theta; \beta}) \operatorname{cosec} \theta / 2\eta_1 \qquad \ldots \quad (6.13)$$

$$k_2 = (d_{\nu_{01}, \nu_{02}; \theta; \alpha} + d_{\nu_{01}, \nu_{02}; \theta; \beta}) \sec \theta / 2\eta_1, \qquad \ldots \quad (6.14)$$

while for the second case they reduce to

$$C = \left[\frac{d_{\nu_{02}, \nu_{02}; \theta; \alpha}}{d_{\nu_{01}, \nu_{02}; \theta; \alpha} + d_{\nu_{01}, \nu_{02}; \theta; \beta}} \right] \eta_1, \qquad \ldots \quad (6.15)$$

$$k_1 = (d_{\nu_{01}, \nu_{02}; \theta; \alpha} + d_{\nu_{01}, \nu_{02}; \theta; \beta}) \operatorname{cosec} \theta / \eta_1, \qquad \ldots \quad (6.16)$$

$$k_2 = (d_{\nu_{01}, \nu_{02}; \theta; \alpha} + d_{\nu_{01}, \nu_{02}; \theta; \beta}) \sec \theta / \eta_1. \qquad \ldots \quad (6.17)$$

Table 5 allows these tests to be applied for $\alpha = 0.025$, $\beta = 0.005$, $n_{01} = 7$ and $n_{02} = 13$. The entries in the rows labelled $(\eta_1 - \eta_0)^2 k_1^2$ and $(\eta_1 - \eta_0)^2 k_2^2$ have been calculated from (6.10) and (6.11), respectively. The italicised entries represent the minimal values of $(\eta_1 - \eta_0)^2 En/\sigma_1^2$, as far as Sukhatme's Tables permit one to judge without interpolation, for the corresponding assumed value of δ. The last two rows have been inserted to enable C to be read off directly for the tests of H_0' and H_0''.

To illustrate the use of Table 5 numerically, suppose we wish to test the hypothesis $\mu_1 \leqslant \mu_2$ against the alternative $\mu_1 > \mu_2$. The requirements to be met are that if $\mu_2 - \mu_2 \leqslant -1$ the probability of rejecting the hypothesis tested must not exceed 0.025, while if $\mu_1 - \mu_2 \geqslant +1$ the risk of 'accepting' the hypothesis tested must not be greater than 0.005. The pilot sample sizes have been chosen to be 7 and 13. There is extraneous evidence available that δ is more likely to be 1 than any other value. The appropriate test is that corresponding to (6.12), (6.13) and (6.14) with $\nu_{01} = 6$, $\nu_{02} = 12$, $\alpha = 0.025$, $\beta = 0.005$, $\eta_0 = -1$, $\eta_1 = +1$. According to Table 5 the average total sample is minimised by choosing $\theta = 45°$ (about). This gives (3rd and 4th rows of the Table) $4k_1^2 = 61.56$, $4k_2^2 = 61.56$. Also, for $\theta = 45°$, $C = -0.1705\eta_1 = -0.1705$ (penultimate row of Table) and the hypothesis is rejected if $\bar{x}_1 - \bar{x}_2 \geqslant -0.1705$,

177

TABLE 5. THE CHARACTERISTIC CONSTANTS FOR THE TWO-STAGE SAMPLE TEST
when $\alpha = 0.025$, $\beta = 0.005$, $n_{01} = 7$, $n_{02} = 13$

	0	15	30	45	60	75	90
$d_{6,\,12;\,\theta;\,0.025}$	2.179	2.193	2.239	2.301	2.368	2.424	2.447
$d_{6,\,12;\,\theta;\,0.005}$	3.055	3.055	3.105	3.247	3.455	3.638	3.707
$(\eta_1-\eta_0)^2 k_1^2$	∞	411.1	114.2	61.56	45.21	39.40	37.87
$(\eta_1-\eta_0)^2 k_2^2$	27.40	29.51	38.08	61.56	135.6	548.6	∞
$(\eta_1-\eta_0)^2 En/\sigma_1^2$ for $\delta=0.01$	∞	411.4	114.6	62.18	46.51	44.89	∞
$(\eta_1-\eta_0)^2 En/\sigma_1^2$ for $\delta=0.1$	∞	414.1	118.0	·67.72	53.77	94.26	∞
$(\eta_1-\eta_0)^2 En/\sigma_1^2$ for $\delta=1$	∞	440.6	152.3	123.1	180.8	588.0	∞
$(\eta_1-\eta_0)^2 En/\sigma_1^2$ for $\delta=5$	∞	558.7	304.6	369.4	723.2	2782	∞
$(\eta_1-\eta_0)^2 En/\sigma_1^2$ for $\delta=10$	∞	706.2	495.0	677.2	1401	5525	∞
$(\eta_1-\eta_0)^2 En/\sigma_1^2$ for $\delta=100$	∞	3362	3922	6217	13610	54900	∞
$\dfrac{d_{\theta;\alpha}-d_{\theta;\beta}}{d_{\theta;\alpha}+d_{\theta;\beta}}$	-0.1673	-0.1643	-0.1620	-0.1705	-0.1863	-0.2020	-0.2017
$\dfrac{d_{\theta;\alpha}}{d_{\theta;\alpha}+d_{\theta;\beta}}$	0.4163	0.4179	0.4190	0.4147	0.4066	0.3999	0.3976

$(d_{\theta;\alpha} \equiv d_{6,12;\theta;0.025}, \; d_{\theta;\beta} \equiv d_{6,12;\theta;0.005})$

If δ is indeed 1, then (7th row) $E(n) = 123.1 \times \sigma_1^2/4 = 30.78\sigma_1^2$. If on the other hand $\delta = 0.1$, then (6th row) $En = 67.72 \times \sigma_1^2/4 = 16.93\sigma_1^2$, and if $\delta = 0.01$, $En = 62.18 \times \sigma_1^2/4 = 15.55\sigma_1^2$, i.e. if $\delta < 1$, the average number of items sampled is actually less than we have estimated. On the other hand, if $\delta > 1$, the reverse is the case, e.g. if $\delta = 5$ instead of $\delta = 1$, then (8th row) $En = 369.4 \times \sigma_1^2/4 = 92.35\sigma_1^2$.

As a second example, consider testing the hypohesis $\mu_1 = \mu_2$ against the alternative $\mu_1 > \mu_2$. The requirements to be met are that if indeed $\mu_1 = \mu_2$ then the risk of rejecting the hypothesis is 0.025, while if it is false and $\mu_1 - \mu_2 \geqslant 1$, then the risk of 'accepting' it must not be greater than 0.005. We suspect that $\delta = 1$, and have (as before) chosen pilot sample sizes 7 and 13. Here, $\nu_{01} = 6$, $\nu_{02} = 12$, $\alpha = 0.025$, $\beta = 0.005$, $\eta_0 = 0$, $\eta_1 = +1$, and the appropriate test is that corresponding to equs. (6.15), (6.16) and (6.17). The minimising value of θ is again approximately 45°, for which $\eta_1^2 k_1^2 = k_1^2 = 61.56$, $\eta_1^2 k_2^2 = k_2^2 = 61.56$. The value of C is 0.4147, and the hypothesis is rejected if $x_1 - x_2 \geqslant 0.4147$.

From (6.2) and (6.9) the limiting power function for one-sided tests of $\eta = \eta_0$ against the set of alternatives $\eta > \eta_0$ on significance level α may be experessed as

$$1 - \psi_{\nu_{01},\,\nu_{02};\,\theta}\left[d_{\nu_{01},\,\nu_{02};\,\theta;\,\alpha} - \frac{\eta - \eta_0}{\left(\dfrac{1}{k_1^2} + \dfrac{1}{k_2^2}\right)^{\frac{1}{2}}} \right]. \qquad \ldots \ (6.18)$$

The value of Q_m corresponding to the region R in the $(\bar{x}_1 \bar{x}_2)$-plane defined by (6.6) is

$$1 - \Phi \left[\frac{\left(\dfrac{1}{k_1^2} + \dfrac{1}{k_2^2} \right)^{\frac{1}{2}} d_{\nu_{01}, \nu_{02}; \theta; \alpha} - (\eta - \eta_0)}{\left(\dfrac{\sigma_1^2}{m_1} + \dfrac{\sigma_2^2}{m_2} \right)^{\frac{1}{2}}} \right], \qquad \ldots \quad (6.19)$$

and the true power function is therefore

$$1 - \sum_{m_1=1}^{\infty} \sum_{m_2=1}^{\infty} p_{m_1}(\sigma_1) p_{m_2}(\sigma_2) \Phi \left[\frac{\left(\dfrac{1}{k_1^2} + \dfrac{1}{k_2^2} \right)^{\frac{1}{2}} d_{\nu_{01}, \nu_{02}; \theta; \alpha} - (\eta - \eta_0)}{\left(\dfrac{\sigma_1^2}{m_1} + \dfrac{\sigma_2^2}{m_2} \right)^{\frac{1}{2}}} \right].$$

$$\ldots \quad (6.20)$$

Q_m is strictly increasing or decreasing in both σ_1 and σ_2 according as to whether $\eta - \eta_0$ is less than or exceeds $\sqrt{(1/k_1^2 + 1/k_2^2)} d_{\nu_{01}, \nu_{02}; \theta; \alpha}$, and is equal to $\frac{1}{2}$ when the latter two quantities are equal, for all σ_1 and σ_2. Consequently, the true power curve lies below or above the limiting power curve according as to whether the first or second inequalities hold, and is hardly affected by variations in the σ_i at the critical value of η. In brief, the test is conservative and is actually somewhat better than stated.

Similarly, to test $\eta = \eta_0$ against $\eta \neq \eta_0$ on significance level α, the hypothesis tested is rejected if

$$\frac{|\bar{x}_1 - \bar{x}_2 - \eta_0|}{\left(\dfrac{1}{k_1^2} + \dfrac{1}{k_2^2} \right)^{\frac{1}{2}}} \geqslant d_{\nu_{01}, \nu_{02}, \theta, \alpha/2} . \qquad \ldots \quad (6.21)$$

This controls the risk of error of the first kind for any pair of chosen values of k_1 and k_2. The limiting power of the test is

$$1 - \psi_{\nu_{01}, \nu_{02}; \theta} \left[d_{\nu_{01}, \nu_{02}; \theta; \alpha/2} - \frac{\eta - \eta_0}{\left(\dfrac{1}{k_1^2} + \dfrac{1}{k_2^2} \right)^{\frac{1}{2}}} \right]$$

$$- \psi_{\nu_{01}, \nu_{02}; \theta} \left[-d_{\nu_{01}, \nu_{02}; \theta; \alpha/2} - \frac{\eta - \eta_0}{\left(\dfrac{1}{k_1^2} + \dfrac{1}{k_2^2} \right)^{\frac{1}{2}}} \right] \qquad \ldots \quad (6.22)$$

Since the density function of the d-statistic is clearly symmetrical round $d = 0$, the limiting power function in (6.22) is strictly increasing in $|\eta - \eta_0|$ and the test is unbiassed.

The procedure for satisfying the additional imposed conditions that the power curve passes through the points $(\eta_0 \pm l, 1 - \beta)$, where β is some given (small) quantity and l is a given positive quantity is straightforward, though rather tedious, to apply in practice. To meet the specifications, we require

$$\beta = \psi_{\nu_{01}, \nu_{02}; \theta} [d_{\nu_{01}, \nu_{02}; \theta; \alpha/2} - l k_1 \sin \theta]$$
$$- \psi_{\nu_{01}, \nu_{02}; \theta} [-d_{\nu_{01}, \nu_{02}; \theta; \alpha/2} - l k_1 \sin \theta]. \qquad \ldots \quad (6.23)$$

The right-hand member of (6.23) may be tabulated or graphed as a function of lk_1, for any chosen value of θ, with the aid of Fisher's extended Tables of the distribution (1941). It equals $1-\alpha$ when $lk_1 = 0$ and decreases steadily to zero as $lk_1 \to \infty$. The value of $hk_1 = lk_1^*$ (say) corresponding to a value β for the function may thus be found, and hence also the appropriate value of $lk_2 = lk_2^* = lk_1^* \tan \theta$. The test is carried through with these values of k_1 and k_2. If desired, this procedure may be repeated for various values of θ, and that value of θ finally chosen which minimises En for some suspected value of $\delta = \sigma_2^2/\sigma_1^2$, as previously.

Finally, the value of Q_m corresponding to (6.21) for the two-sided test is

$$1 - \Phi \left[\frac{(1/k_1^2 + 1/k_2^2)^{\frac{1}{2}} d_{\nu_{01}, \nu_{02}; \theta; \alpha/2} - (\eta - \eta_0)}{(\sigma_1^2/m_1 + \sigma_2^2/m_2)^{\frac{1}{2}}} \right]$$

$$- \Phi \left[\frac{-(1/k_1^2 + 1/k_2^2)^{\frac{1}{2}} d_{\nu_{01}, \nu_{02}; \theta; \alpha, 2} - (\eta - \eta_0)}{(\sigma_1^2/m_1 + \sigma_2^2/m_2)^{\frac{1}{2}}} \right], \qquad \dots \quad (6.24)$$

and the true power function is

$$1 - \sum_{m_1=1}^{\infty} \sum_{m_2=1}^{\infty} p_{m_1}(\sigma_1) \, p_{m_2}(\sigma_2) \left\{ \Phi \left[\frac{(1/k_1^2 + 1/k_2^2)^{\frac{1}{2}} d_{\nu_{01}, \nu_{02}; \theta; \alpha/2} - (\eta - \eta_0)}{(\sigma_1^2/m_1 + \sigma_2^2/(m_2)^{\frac{1}{2}}} \right] \right.$$

$$\left. - \Phi \left[\frac{-(1/k_1^2 + 1/k_2^2)^{\frac{1}{2}} d_{\nu_{01}, \nu_{02}; \theta; \alpha/2} - (\eta - \eta_0)}{(\sigma_1^2/m_1 + \sigma_2^2/m_2)^{\frac{1}{2}}} \right] \right\}. \qquad \dots \quad (6.25)$$

As for the one-sided test, inspection of Q_m shows that the test is conservative.

REFERENCES

BARTLETT, M. S. (1936) : The information available in small samples. *Proc. Camb. Phil. Soc.*, 32, 560–566.

———— (1937) : Properties of sufficiency and statistical tests. *Proc. Roy. Soc.*, A, **160**, 268-282.

———— (1956) : Comment on Sir Ronald Fisher's paper : "On a test of significance in Pearson's Biometrika Tables (No. 11)". *J. Roy. Stat. Soc.*, B, **18**, 295-296.

BEHRENS, W. V. (1929) : Ein Betrag zur Fehlerberechnung bei wenigen Beobachtungen. *Landw. Jb.*, **68**, 807–837.

CHAPMAN, D. G. (1950) : Some two sample tests. *Ann. Math. Stat.*, **21**, 601-606.

FISHER, R. A. (1935) : The fiducial argument in statistical inference,. *Ann. Eugen.*, **6**, 391–398.

———— (1941) : The asymptotic approach to Behrens' integral, with further tables for the *d*-test of significance. *Ann. Eugen.*, **11**, 141-173.

———— (1956) : *Statistical Methods and Scientific Inference*, Oliver and Boyd, Edinburgh.

FISHER, R. A. and YATES, F. (1957) : *Statistical Tables for Biological, Agricultural and Medical Research*, Oliver and Boyd, Edinburgh.

RUBEN, H. (1950) : *Sequential Studentisation of Sampling Means from Unknown Normal Populations.* Ph.D. dissertation, London University.

———— (1960) : On the distribution of the weighted difference of two independent Student variables. *J. Roy. Stat. Soc.*, B, **22**, 188-194.

———— (1961) : Studentisation of two-stage sampling means from normal populations with unknown common variance. *Sankhyā*, Series A, **23**, 231–250.

STEIN, C. (1945) : A two-sample test for a linear hypothesis whose power is independent of the variance. *Ann. Math. Stat.*, **16**, 243-258.

SUKHATME, P. V. (1938) : On Fisher and Behrens test of significance for the difference in means of two normal samples. *Sankhyā*, **4**, 39-48.

Paper received : December, 1960.

A LAGUERRE SERIES APPROXIMATION TO THE SAMPLING DISTRIBUTION OF THE VARIANCE

By J. ROY and M. L. TIKU

Indian Statistical Institute

SUMMARY. The first four terms of the Laguerre series expansion of the distribution of the variance of a sample from any population are worked out.

1. INTRODUCTION

An approximation to the sampling distribution of the variance in samples from any non-normal population was given by Gayen (1949). He started with a Gram-Charlier series expansion of the probability density function of the population and ignored all cumulants of the population above the fourth and also squares and higher powers of the fourth cumulant. An alternative approach is presented in this paper. The probability density function of the sample variance is expanded in terms of a Gamma density function and Laguerre polynomials and the coefficients of the first four terms are worked out explicitly in terms of population cumulants, of upto the eighth order. Gayen's expression agrees with this expansion to the order of approximation used by Gayen.

Laguerre polynomials. For $m > 0$, a Laguerre polynomial of degree r in x is defined as (see Szegö (1939), Chapter V)

$$L_r^{(m)}(x) = \sum_{t=0}^{r} C_{r,t}^{(m)} \frac{(-x)^t}{t!}, \qquad \dots (1.1)$$

where
$$C_{r,t}^{(m)} = \begin{cases} (m+t)(m+t+1) \dots (m+r-1)/(r-t)!, & \text{for } t = 0, 1, \dots, r-1 \\ 1 & \text{for } t = r \end{cases}$$

for $r = 0, 1, 2, \dots$. In particular, the first four polynomials are:

$$L_0^{(m)}(x) = 1$$

$$L_1^{(m)}(x) = m - x$$

$$L_2^{(m)}(x) = \frac{1}{2!} \; m(m+1) - (m+1)x + \frac{x^2}{2!} \qquad \dots (1.2)$$

$$L_3^{(m)}(x) = \frac{1}{3!} m(m+1)(m+2) - \frac{1}{2!}(m+1)(m+2)x + (m+2)\frac{x^2}{2!} - \frac{x^3}{3!}$$

$$L_4^{(m)}(x) = \frac{1}{4!} \; m(m+1)(m+2)(m+3) - \frac{1}{3!}(m+1)(m+2)(m+3)x$$

$$+ \frac{1}{2!}(m+2)(m+3)\frac{x^2}{2!} - (m+3)\frac{x^3}{3!} + \frac{x^4}{4!}.$$

If we write

$$p_m(x) = \frac{1}{\Gamma(m)} e^{-x} x^{m-1} \qquad \qquad \dots \quad (1.3)$$

for the Gamma density function with mean m, the orthogonality property of Laguerre polynomials can be stated as :

$$\int_0^\infty L_r^{(m)}(x) L_s^{(m)}(x) p_m(x) dx = \begin{cases} 0 & \text{if } r \neq s \\ \\ C_{r,0}^{(m)} & \text{if } r = s \end{cases} \qquad \dots \quad (1.4)$$

2. Approximate distribution of the corrected sum of squares

Let $Y_1, Y_2, \dots, Y_n$ be a random sample from a population in which cumulants of all orders exist, and let the r-th cumulant be denoted by $K_r, r = 2, 3, \dots$. The variance of the population will be alternatively denoted by $\sigma^2 = K_2$. Further, we shall write

$$\lambda_r = K_r K_2^{-\frac{1}{2}r} ; \quad r = 3, 4, \quad \dots \qquad \dots \quad (2.1)$$

Consider the sum of squares about the sample mean : $S^2 = \sum_{i=1}^n (Y_i - \bar{Y})^2$ where $\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$.
This is said to have $\nu = (n-1)$ degrees of freedom. We shall write

$$X = S^2/2\sigma^2 \qquad \dots \quad (2.2)$$

and try to derive an approximation for the probability density function of X.

We first note that the cumulants of the distribution of $s^2 = S^2/\nu$ has been worked out by Fisher (1928), and that the first few are tabulated in Kendall (1947). Using these, we get the first four central moments of X and these are listed below :

$$E(X) = m$$

$$\mu_2(X) = m + \frac{m^2}{2m+1} \lambda_4$$

$$\mu_3(X) = 2m + \frac{6m^2}{2m+1} \lambda_4 + \frac{m(2m-1)}{2m+1} \lambda_3^2 + \frac{m^3}{(2m+1)^2} \lambda_6 \qquad \dots \quad (2.3)$$

$$\mu_4(X) = 3m(m+2) + \frac{12m^2(m+3)}{2m+1} \lambda_4 + \frac{12m(2m-1)}{2m+1} \lambda_3^2 + \frac{12m^3}{(2m+1)^2} \lambda_6$$

$$+ \frac{m(3m^3+16m^2-2m+1)}{(2m+1)^2} \lambda_4^2 + \frac{8m^2(2m-1)}{(2m+1)^2} \lambda_5 \lambda_3 + \frac{m^4}{(2m+1)^3} \lambda_8$$

where we write for simplicity $\quad m = \frac{1}{2}(n-1)$. $\qquad \dots \quad (2.4)$

Let us denote by $\phi_m(x)$ the probability density function of X. The quotient $\phi_m(x)/p_m(x)$ can be formally expanded in an infinite series in Laguerre polynomials as

$$\frac{\phi_m(x)}{p_m(x)} = \sum_{r=0}^\infty a_r^{(m)} L_r^{(m)}(x). \qquad \dots \quad (2.5)$$

Multiplying both sides of (2.5) by $L_r^{(m)}(x)\, p_m(x)$ and integrating over x from 0 to ∞, we get

$$a_r^{(m)} = \int_0^\infty L_r^{(m)}(x)\phi_m(x)dx/C_{r,0}^{(m)} \qquad \dots \ (2.6)$$

formally, by virtue of the orthogonal property (1.4) of Laguerre polynomials. For conditions of convergence of the formal expansion (2.5) see Szegö (1939), Chapter IX.

What we seek here is an approximation to $\phi_m(x)$ using only the first four terms in (2.5). Thus

$$\phi_m(x) \sim p_m(x) \left[\sum_{r=0}^{4} a_r^{(m)}\, L_r^{(m)}(x) \right] \qquad \dots \ (2.7)$$

where

$$a_r^{(m)} = \frac{E[L_r^{(m)}(X)]}{C_{r,0}^{(m)}}, \ r = 0, 1, 2, 3, 4. \qquad \dots \ (2.8)$$

To evaluate $E[L_r^{(m)}(X)]$ for $r = 0, 1, 2, 3$ and 4, we note that writing $y = x - m$, the first four Laguerre polynomials given by (1.2) can be expressed in terms of y as :

$$L_1^{(m)} = -y$$

$$L_2^{(m)} = \frac{1}{2}y^2 - y - \frac{1}{2}m \qquad \qquad \dots \ (2.9)$$

$$L_3^{(m)} = -\frac{1}{6}y^3 + y^2 + \left(\frac{1}{2}m - 1\right)y - \frac{2}{3}m$$

$$L_4^{(m)} = \frac{1}{24}y^4 - \frac{1}{2}y^3 + \frac{1}{4}(6-m)y^2 + \left(\frac{7}{6}m - 1\right)y + \frac{1}{8}m(m-6).$$

Hence

$$E[L_1^{(m)}(X)] = 0$$

$$E[L_2^{(m)}(X)] = \frac{1}{2}\mu_3 - \frac{1}{2}m \qquad \qquad \dots \ (2.10)$$

$$E[L_3^{(m)}(X)] = -\frac{1}{6}\mu_3 + \mu_2 - \frac{2}{3}m$$

$$E[L_4^{(m)}(X)] = \frac{1}{24}\mu_4 - \frac{1}{2}\mu_3 + \frac{1}{4}(6-m)\mu_2 + \frac{1}{8}m(m-6)$$

where μ_2, μ_3, μ_4 are respectively the second, third and fourth central moments of X given by (2.3). Using (2.8) and (2.10), we finally get

$$a_0^{(m)} = 1$$
$$a_1^{(m)} = 0$$
$$a_2^{(m)} = \frac{m}{(2m+1)(m+1)}\ \lambda_4 \qquad \qquad \dots \ (2.11)$$

$$a_3^{(m)} = -\frac{1}{(2m+1)(m+1)(m+2)}\left[\frac{m^3}{2m+1}\lambda_6 + (2m-1)\ \lambda_4^2 \right]$$

$$a_4^{(m)} = \frac{1}{(2m+1)(m+1)(m+2)(m+3)}\left[\frac{m^3}{(2m+1)^2}\lambda_8 + \frac{8m(2m-1)}{2m+1}\ \lambda_5\lambda_3 \right.$$
$$\left. + \frac{3m^3+16m^2-2m+1}{(2m+1)}\ \lambda_4^2 \right]$$

The probability density function of the sample variance $Z = \dfrac{S^2}{n-1} = \dfrac{\sigma^2 X}{m}$ is then obtained as

$$m\sigma^{-2}p_m(m\sigma^{-2}z) \sum_{r=0}^{4} a_r^{(m)} L_r^{(m)}(m\sigma^{-2}z) \qquad \dots \quad (2.12)$$

If terms involving λ_r for $r > 4$ and λ_4^2 are ignored, this agrees with Gayen's (1949) formula (3.1).

References

GAYEN, A. K. (1949) : The distribution of Student's t in random samples of any size from non-normal universes. *Biometrika*, **36**, 353-369.

SZEGÖ, G. (1939) : Orthogonal polynomials (*Amer. Math. Soc. Colloquium Publication* Vol. XXIII).

KENDALL, M. G. (1947) : *Advanced Theory of Statistics*, Vol. I, Third edition, Charles Griffin & Company Ltd.

FISHER, R. A. (1928) : Moments and product moments of sampling distributions. *Proc. Lond. Math. Soc.*, **30**, 199-238.

Paper received : December, 1961.

MAIN EFFECTS AND INTERACTIONS*

By HENRY B. MANN

Ohio State University

SUMMARY. Expected yields of observations, for an arbitrary region of factor space are decomposed into main effects and interactions of the various factors present. Methods to test the significance of the contribution to the highest order interaction from a sub-set of the factor space are developed for the cases when (1) some *a priori* information about the interaction is available, (2) there is no *a priori* information. The test used here is shown to be more powerful than the test for the highest order interaction in the entire region.

1. INTRODUCTION

In a factorial experiment on two factors we apply each factor on varying levels to various experimental units. We shall assume that this application yields for each unit a quantity which we shall call the yield of this unit. We denote by $f(x, y)$ the mean value of the yield obtained, when the first factor is applied at level x and the second factor at level y. The levels may be capable of discrete values only or they may vary continuously.

Intuitively one feels that the function $f(x, y)$ should be broken up into a general mean μ, an effect $g(x)$ of the first factor, an effect $h(y)$ of the second factor and an effect $u(x, y)$ ascribed to the combination of level x of the first factor with level y of the second factor. The function $g(x)$ is to denote the deviation from the general mean due to the application of the level x of the first factor. The function $u(x, y)$ is to denote a deviation from the mean value of $f(x, y) - g(x)$ when y is fixed. Accordingly we should like to impose the restrictions

$$\int g(x)d(x) = \int h(y)dy = \int u(x, y)dx = \int u(x, y)dy = 0, \qquad \ldots (1.1)$$

where the integrals have to be replaced by sums if the levels are discrete and throughout this paper each integral (or sum) is extended over the domain of definition of the integrand (or summand) unless otherwise stated. We then have the decomposition

$$f(x, y) = \mu + g(x) + h(y) + u(x, y). \qquad \ldots (1.2)$$

The quantities $g(x)$, $h(y)$, μ and $u(x, y)$ are estimated by least squares. This amounts to defining $g(x)$ and $h(y)$ by the condition that

$$\iint (f(x, y) - \mu - g(x) - h(y))^2 dydx \qquad \ldots (1.3)$$

should be minimized by the choice of the quantities μ, $g(x)$, $h(y)$. Differentiation with respect to μ and with respect to $g(x)$ on each y level and $h(y)$ on each x level (or alternately a simple application of the calculus of variations) leads to

$$\iint (f(x,y)-\mu-g(x)-h(y))dydx = 0,$$
$$\int (f(x,y)-\mu-g(x)-h(y))dx = 0,$$
$$\int (f(x,y)-\mu-g(x)-h(y))dy = 0. \qquad \dots \quad (1.4)$$

Now assume that $f(x,y)$ is defined for $0 \leqslant x \leqslant a$, $0 \leqslant y \leqslant b$. Then (1.4) and (1.1) lead to the unique solution

$$\mu = \frac{1}{ab} \int_0^a \int_0^b f(x,y)dxdy, \; g(x) = \frac{1}{b} \int_0^b f(x,y)dy - \mu,$$

$$h(y) = \frac{1}{a} \int_0^a f(x,y)dx - \mu. \qquad \dots \quad (1.5)$$

If the levels are discrete the solutions are obtained by replacing integrals by sums.

In the following we shall restrict ourselves to the case that the levels are discrete, but many of the statements and proofs can easily be modified to apply to continuous levels.

The idea can easily be generalized to any number of factors. We introduce the decomposition

$$f(a_1, \dots, a_n) = \sum_{\alpha=0}^{n} \sum_{1,\dots,n} \mu_{a_{i_1}, \dots, a_{i_\alpha}}^{i_1, \dots, i_\alpha}, \qquad \dots \quad (1.6)$$

where $\sum_{1,\dots,n}$ means summation over all choices $i_1 < \dots < i_\alpha$ out of $1, \dots, n$ and the quantities $\mu_{a_{i_1}, \dots, a_{i_\alpha}}^{i_1, \dots, i_\alpha}$ satisfy the conditions

$$\sum_{a_\beta} \mu_{a_1, \dots, a_\alpha}^{i_1, \dots, i_\alpha} = 0, \qquad \beta = 1, \dots, \alpha. \qquad \dots \quad (1.7)$$

(In (1.7) all a's except a_β are fixed and a_β runs over all values for which the summand is defined.) That the decomposition (1.6) is always possible will follow from Theorem 1.

The quantity μ_a^i is called the main effect of the a-th level of the i-th factor. The quantity $\mu_{a_1, \dots, a_\alpha}^{i_1, \dots, i_\alpha}$, $\alpha > 1$, is called the interaction between the levels $a_1, \dots, a_\alpha$ of the factors $i_1, \dots, i_\alpha$. The totality of all $\mu_{a_1, \dots, a_\alpha}^{i_1, \dots, i_\alpha}$ for all possible values $a_1, \dots, a_\alpha$ is called the interaction $(i_1, \dots, i_\alpha)$.

186

A set I of interactions will be called a natural set of interactions if $(i_1, \ldots, i_\alpha)\epsilon I$, $(i_1, \ldots, i_\alpha)\epsilon(j_1, \ldots, j_\beta)$ implies $(j_1, \ldots, j_\beta)\epsilon I$.

Theorem 1 : *Let S be any set of points $(a_1, \ldots, a_n)$, let $x_{a_1, \ldots, a_n}$ be a function defined on S and let I be a natural set of interactions. If the equations*

$$x_{a_1, \ldots, a_n} = \sum_I m_{a_{i_1}, \ldots, a_{i_\alpha}}^{i_1, \ldots, i_\alpha} \qquad \ldots \ (1.8)$$

have a solution for $m_{a_1, \ldots, a_\alpha}^{i_1, \ldots, i_\alpha}$ then there also exists a solution $\mu_{a_1, \ldots, a_\alpha}^{i_1, \ldots, i_\alpha}$ satisfying (1.7).

The solution $\mu_{a_1, \ldots, a_\alpha}^{i_1, \ldots, i_\alpha}$ is however not necessarily unique.

Proof : For $\alpha = 0$ in (1.7) there is nothing to prove. By induction we may assume that in (1.8) all $m_{a_1, \ldots, a_\alpha}^{i_1, \ldots, i_\alpha}$ satisfy (1.7) for $\alpha < u$. For simplicity assume $1, \ldots, u \epsilon I$. Consider the quadratic form

$$Q = \sum_{a_1} \ldots \sum_{a_u} \left(m_{a_1, \ldots, a_u}^{1, \ldots, u} - \sum_{\alpha=0}^{u-1} \sum_{1, \ldots, u} V_{a_{i_1} \ldots a_{i_\alpha}}^{i_1, \ldots, i_\alpha} \right)^2 . \qquad \ldots \ (1.9)$$

If we minimize Q with respect to the parameters $V_{a_{i_1} \ldots, a_{i_\alpha}}^{i_1, \ldots, i_\alpha}$ we get among others the equations

$$\sum_{b_i} \left(m_{a_1, \ldots, a_{i-1} \ b_i \ a_{i+1}, \ldots, a_u}^{1, \ldots, i-1 \ i \ i+1, \ldots, u} - \sum_{\alpha=0}^{u-1} \sum_{1, \ldots, u} V_{a_{i_1}, \ldots, a_{i_\alpha}}^{i_1, \ldots, i_\alpha} \right) = 0 \qquad \ldots \ (1.10)$$

resulting from differentiation of Q with respect to $V_{a_1, \ldots, a_{i-1} \ a_{i+1}, \ldots, a_u}^{1, \ldots, i-1 \ i+1, \ldots, u}$. Since Q has a minimum (1.10) must have a solution. By induction we may assume that the solutions $V_{a_{i_1}, \ldots, a_{i_\alpha}}^{i_1, \ldots, i_\alpha}$ satisfy (1.7). We now put

$$m*_{a_1, \ldots, a_u}^{1, \ldots, u} = m_{a_1, \ldots, a_u}^{1, \ldots, u} - \sum_{\alpha=0}^{u-1} \sum_{1, \ldots, u} V_{a_{i_1}, \ldots, a_{i_\alpha}}^{i_1, \ldots, i_\alpha} ,$$

$$m*_{a_1, \ldots, a_\alpha}^{i_1, \ldots, i_\alpha} = m_{a_1, \ldots, a_\alpha}^{i_1, \ldots, i_\alpha} + V_{a_1, \ldots, a_\alpha}^{i_1, \ldots, i_\alpha} \quad \begin{array}{l} \text{for } \alpha < u, \\ i_\alpha \leqslant u \end{array} \qquad \ldots \ (1.11)$$

$$m*_{a_1, \ldots, a_\alpha}^{i_1, \ldots, i_\alpha} = m_{a_1, \ldots, a_\alpha}^{i_1, \ldots, i_\alpha} \qquad \text{for } i_\alpha > u.$$

Clearly $\qquad x_{a_1, \ldots, a_n} = \sum_I m*_{a_{i_1}, \ldots, a_i}^{i_1, \ldots, i_\alpha} ;$

where $m*_{a_1, \ldots, a_u}^{1, \ldots, u}$ and $m*_{a_1, \ldots, a_\alpha}^{i_1, \ldots, i_\alpha}, \alpha < u$, satisfy (1.7). We can apply the same process for any set $i_1, \ldots, i_u$ in I. Hence the theorem.

187

Corollary 1 : *The representation (1.6) is possible for any set S and any function $X_{a_1, ..., a_n}$ defined in S.*

For we can simply put $m_{a_1, ..., a_n} = X_{a_1, ..., a_n}$, $m^{i_1, ..., i_\alpha}_{a_{j_1}, ..., a_{j_\alpha}} = 0$ for $\alpha < n$ and apply Theorem 1.

If $X_{a_1, ..., a_n}$ is defined in the region $T = \{1 \leqslant a_1 \leqslant t_1, ..., 1 \leqslant a_n \leqslant t_n\}$ then (1.6) and (1.7) have a unique solution

$$\mu^{i_1, ..., i_\alpha}_{a_{i_1}, ..., a_{i_\alpha}} = \sum_{\beta=0}^{\alpha} (-1)^{\alpha-\beta} \sum_{i_1, ..., i_\alpha} X^{k_1, ..., k_\beta}_{a_{k_1}, ..., a_{k_\beta}} , \qquad \ldots \quad (1.12)$$

where $X^{k_1, ..., k_\beta}_{a_1, ..., a_\beta}$ is the mean of all values $X_{b_1, ..., b_n}$ with $b_{k_1} = a_1, ..., b_{k_\beta} = a_\beta$.

For the derivation of this formula and more detailed discussions, see Mann (1949).

The main effects and interactions change if the region changes in which the function $f(x_1, ..., x_n)$ is considered. In a given experiment this region depends on our choice of levels of the factors and we shall therefore call it the design of the experiment. We can give the main effects and interactions an absolute meaning if we consider a fixed region T and consider only designs from this region. One might contemplate procedures for choosing a design or a sample of designs which would permit unbiased estimation of the "true" main effects and interactions (i.e. those arising from the region T). However, this is not the goal of the present investigation. The fact that the main effects and interactions depend on the design seems to detract from the value of factorial experiments and of the decomposition (1.6). However, there is often considerable interest in the values which the main effects and interactions take in a particular design irrespective of their "true" values. There is moreover a great deal of valuable information to be extracted from the way in which the interactions change when the design is modified.

To illustrate this point consider a nutritional experiment in which various levels of carbohydrate and of protein are fed to animals and their weight gains after some period are recorded. We assume that all carbohydrate levels administered include sufficient amounts to supply the energy requirements of the animals. Let us make the simplifying, but approximately correct, assumption that there is a certain amount p of protein necessary for the proper growth and upkeep of the animal and that any additional amounts of protein will be utilized in the same way as carbohydrates are used. Thus amounts of protein exceeding the level p can be substituted by proper amounts of carbohydrates. On the other hand, if the protein level is less than p we assume that increasing the carbohydrate level will have no effect on the animal. (This may be an oversimplification, but it is not the author's intention to write a paper on animal nutrition.) Under these assumptions, if all protein levels administered exceed the minimum level p, we shall find no interaction, because protein and carbohydrate can be substituted for each other. On the other hand, if all protein

levels are below p, then the weight of the animal will depend only on the amount of protein fed, the carbohydrate effect will be zero and again the interaction will be zero. We shall find an interaction only if the protein levels include levels above and below p. Thus the dependence of the interaction on the design reflects facts which are of basic importance in animal nutrition. On the other hand, if we carry out an experiment including high and low protein levels and compute the interaction from this experiment, we may find high interactions in all experimental units and it may not be possible just by "looking at the data" to find that the interaction arose from the inclusion of the low levels. It seems therefore of importance to develop procedures which will help us to locate the levels whose inclusion gives rise to the interaction.

2. TESTS FOR INTERACTIONS

As in many other applications of the analysis of variance, two situations have to be distinguished. We may (Case I) feel justified in assuming that interaction, if it arises at all, will arise from the inclusion of certain specified levels, known in advance; or we may (Case II) not be in possession of such *a priori* information. We shall develop methods for each case. The computational problems in both cases are similar, but the applications and interpretation of our statements are different.

Case I fits completely into the scheme of testing hypotheses. Let T be the total range of our experiment and let S be the subrange suspected of giving rise to the highest order interaction. Our Assumption A then is that all highest order interactions are 0 in $T - S$ and we test the hypothesis that these interactions are 0 in T. (It follows from Theorem 1 that the absence of highest order interactions in T implies absence of such interactions in $T - S$.) In formal language our linear hypothesis reads :

$$A : E(x_{a_1, \ldots, a_n}) = \sum_{\alpha=0}^{n-1} \sum_{1, \ldots, n} \mu \begin{smallmatrix} i_1, \ldots, i_\alpha \\ a_{i_1}, \ldots, a_{i_\alpha} \end{smallmatrix} \quad \text{in} \quad T-S, \qquad \ldots \quad (2.1)$$

$$H : E(x_{a_1, \ldots, a_n}) = \sum_{\alpha=0}^{n-1} \sum_{1, \ldots, n} \bar{\mu} \begin{smallmatrix} i_1, \ldots, i_\alpha \\ a_{i_1}, \ldots, a_{i_\alpha} \end{smallmatrix} \quad \text{in} \quad T.$$

The test proceeds along the lines given by Mann (1949, pp. 23 ff). The rank of the matrix (g_{i_α}) of eqn. 4.6 of Mann (1949) must however be found first and this is not always easy.

The procedure is as follows : Consider a matrix whose rows are numbered by all combinations $(a_1, \ldots, a_n)$ occurring in $T - S$ and whose columns are headed by all symbols $\begin{smallmatrix} i_1, \ldots, i_\alpha \\ b_1, \ldots, b_\alpha \end{smallmatrix}$, $\alpha \leqslant n-1$, such that at least one experiment occurs in $T - S$ for which the factors $i_1, \ldots, i_\alpha$ are at levels $b_1, \ldots, b_\alpha$. Write 1 into the intersection of row $a_1, \ldots, a_n$ and column $\begin{smallmatrix} i_1, \ldots, i_\alpha \\ b_1, \ldots, b_\alpha \end{smallmatrix}$ if $a_{i_1} = b_1, \ldots, a_{i_\alpha} = b_\alpha$, otherwise write 0. Let p be the rank of this matrix and N the number of observations in $T - S$. Assume $N > p$ (otherwise no test is possible, unless an independent estimate of the variance is available).

Minimize $Q = \sum_{a_1} \ldots \sum_{a_n} (x_{a_1, \ldots, a_n} - E(x_{a_1, \ldots, a_n}))^2$ under the assumptions A, where the sum is extended over all combinations $a_1, \ldots, a_n$ occurring in T. Write Q_a for the minimum value of Q under A. Since $E(x_{a_1, \ldots, a_n})$ is unrestricted in S, Q_a can be found by minimizing the sum of squares extended over the combinations in $T-S$ only.

Find the rank s of a similar matrix for the space T and let Q_r be the minimum of Q under the hypothesis H. Let N_1 be the number of observations in T.

The assumption A imposes $N-p$ independent linear restrictions so that Q_a has $N-p$ degrees of freedom. The hypothesis H imposes N_1-s independent linear restrictions, hence $N_1-N-s+p$ additional linear restrictions. Hence, if $x_{a_1, \ldots, a_n}$ are normally and independently distributed with the same variance and if H is true, the statistic

$$F = \frac{N-p}{N_1-N+p-s} \frac{Q_r-Q_a}{Q_a} \qquad \ldots \quad (2.2)$$

has the F distribution with $N_1-N+p-s$ and $N-p$ degrees of freedom.

In experiments where more than one but an equal number of observations r are taken in each subclass one can obtain an estimate $\hat{\sigma}^2$ of the variance from the sums of deviations of each observation from its cell mean and this estimate has $(r-1)N_1$, degrees of freedom. The quantities Q_r-Q_a must then be computed from $\sqrt{r} \, x_{a_1, \ldots, a_n}$, where the $x_{a_1, \ldots, a_n}$ are the cell means.

To test the hypothesis H under the assumption A one then should use

$$F = \frac{N-p+(r-1)N_1}{N_1-N+p-s} \frac{Q_r-Q_a}{(r-1)N_1\hat{\sigma}^2+Q_a} \qquad \ldots \quad (2.3)$$

In this case or any other case where an independent estimate $\hat{\sigma}^2$ with say V degrees of freedom is available one can also test the contribution of the region S to the highest order interaction sum of squares by using

$$F = \frac{V}{N_1-N+p-s} \frac{Q_r-Q_a}{V\hat{\sigma}^2} \qquad \ldots \quad (2.4)$$

The statistic (2.4) will be more powerful than a test for the highest order interaction in T, if these interactions are low in $T-S$ compared to the interactions in T, particularly so if $T-S$ is large compared to S.

The procedure is similar if the assumption consists of the statement that certain interaction sets I_1 are 0 in T and some additional interaction sets I_2 are 0 in $T-S$ while H states that the interactions of $I_1 \bigcup I_2$ are 0 in T. It is of course necessary that H really implies A. It follows from Theorem 1 that this is always the case if the complement of $I_1 \bigcup I_2$ is a natural set of interactions.

3. Tests when *a priori* information is available

The actual computation of Q_r and Q_a for arbitrary regions T, S may become very involved. Usually, however, the regions T and S are the integral points of rectangular parallelopipeds. Without loss of generality we may in this case assume

$$T = \{a_1, ..., a_n;\ 1 \leqslant a_i \leqslant t_i\},\ S = \{a_1, ..., a_n;\ 1 \leqslant a_i \leqslant s_i\}.$$

We shall derive Q_a for the linear hypothesis (2.1) in this case. The value of Q_r is well known and is derived by Mann (1949, eqns. 5.5 and 5.9). To find Q_a we prove the following lemma.

Lemma 1: *Let S be any region and let $x_{a_1, ..., a_n}$ be defined over S. Let*

$$x_{a_1, ..., a_n} - \sum_{0}^{n} \sum_{1, ..., n} \mu^{i_1, ..., i_\alpha}_{a_{i_1}, ..., a_{i_\alpha}} = 0, \qquad \qquad ...\ (3.1)$$

where the $\mu^{i_1, ..., i_\alpha}_{a_1, ..., a_\alpha}$ satisfy (1.7). Then the values $\mu^{i_1, ..., i_\alpha}_{a_1, ..., a_\alpha}$, $\alpha < n$, satisfy the least square equations obtained under the assumptions that $\mu^{1, ..., n}_{a_1, ..., a_n} = 0$ for all $a_1, ..., a_n\ \epsilon S$.

Proof: If we take the derivative of

$$Q = \sum_{a_1} ... \sum_{a_n} \left(x_{a_1, ..., a_n} - \sum_{\alpha=0}^{n-1} \sum_{1, ..., n} \mu^{i_1, ..., i_\alpha}_{a_{i_1}, ..., a_{i_\alpha}} \right)^2$$

with respect to $\mu^{1, ..., \alpha}_{a_1, ..., a_\alpha}$, $\alpha < n$, we get the least square equations

$$\sum_{a_{\alpha+1}} ... \sum_{a_n} \left(x_{a_1, ..., a_n} - \sum_{\alpha=0}^{n-1} \sum_{1, ..., n} \mu^{i_1, ..., i_\alpha}_{a_{i_1}, ..., a_{i_\alpha}} \right) = 0. \qquad ...\ (3.2)$$

Because of (1.7), however, the same equations are obtained by extending the sum in (3.2) over the equations (3.1). This proves the lemma.

To obtain the least square estimates $A^{i_1, ..., i_\alpha}_{a_1, ..., a_\alpha}$ for $\mu^{i_1, ..., i_\alpha}_{a_1, ..., a_\alpha}$ in $T - S$ we may therefore set

$$x_{a_1, ..., a_n} = \sum_{\alpha=0}^{n} \sum_{1, ..., n} A^{i_1, ..., i_\alpha}_{a_{i_1}, ..., a_{i_\alpha}} \quad \text{in}\ T - S \qquad ...\ (3.1')$$

and compute all $A^{i_1, ..., i_\alpha}_{a_1, ..., a_\alpha}$ from (3.1'). We then have

$$Q_a = \sum_{a_1} ... \sum_{a_n} \left(A^{1, ..., n}_{a_1, ..., a_n} \right)^2. \qquad \qquad ...\ (3.3)$$

To carry out the calculations we set

$$y_{a_1, ..., a_n} = \sum_{\alpha=0}^{n} \sum_{1, ..., n} A^{i_1, ..., i_\alpha}_{a_{i_1}, ..., a_{i_\alpha}} \quad \text{in}\ T, \qquad ...\ (3.4)$$

191

12

where we substitute on the right the values of (3.1′) if the combination $a_{i_1}, ..., a_{i_\alpha}$ occurs in $T-S$ and 0 otherwise. Clearly, any set of values which solves (3.4) will also solve (3.1′).

The equations (3.4) have the unique solution [Mann (1949) eqn. 5.5].

$$A^{i_1, ..., i_\alpha}_{a_{i_1}, ..., a_{i_\alpha}} = \sum_{\beta=0}^{\alpha} (-1)^{\beta-\alpha} \sum_{i_1, ..., i_\alpha} y^{k_1, ..., k_\beta}_{a_{k_1}, ..., a_{k_\beta}} . \qquad ... (3.5)$$

Now let $b_1 \leqslant s_1, ..., b_u \leqslant s_u$ be fixed, $s_{u+1} < t_{u+1}$, $s_l < t_l$, $s_{l+1} = t_{l+1}, ..., s_n = t_n$. (We permit any rearrangement of the indices and it is permissible that $s_v = t_v$ for $v \leqslant u$.) Then if $a_j > s_j$ for at least one value of a_j we have

$$y^{1, ..., u, u+1, ..., l}_{b_1, ..., b_u, a_{u+1}, ..., a_l} = x^{1, ..., u, u+1, ..., l}_{b_1, ..., b_u, a_{u+1}, ..., a_l} \qquad ... (3.6)$$

and therefore from (3.5)

$$x^{1, ..., u, u+1, ..., l}_{b_1, ..., b_u, a_{u+1}, ..., a_l} - A^{1, ..., u, u+1, ..., l}_{b_1, ..., b_u, a_{u+1}, ..., a_l}$$

$$= \sum_{\beta=0}^{u} \sum_{\alpha=0}^{l-u-1} (-1)^{l-\alpha-\beta+1} \sum_{1, ..., u} \sum_{u+1, ..., l} y^{i_1, ..., i_\beta, j_1, ..., j_\alpha}_{b_{i_1}, ..., b_{i_\beta}, a_{j_1}, ..., a_{j_\alpha}}, \qquad ... (3.7)$$

where the second double sum is extended over all choices $i_1 < ... < i_\beta, j_1 < ... < j_\alpha$ where $i_1, ..., i_\beta$ are chosen from $1, ..., u$ and $j_1, ..., j_\alpha$ from $u+1, ..., l$. On account of (3.6) if $a_j > s_j$ for all $u < j \leqslant l$ this may be written

$$x^{1, ..., u, u+1, ..., l}_{b_1, ..., b_u, a_{u+1}, ..., a_l} - A^{1, ..., u, u+1, ..., l}_{b_1, ..., b_u, a_{u+1}, ..., a_l} = (-1)^{l-u-1} \sum_{\beta=0}^{u} (-1)^{u-\beta} \sum_{1, ..., u} y^{i_1, ..., i_\beta}_{b_{i_1}, ..., b_{i_\beta}}$$

$$+ \sum_{\beta=0}^{u} \sum_{\alpha=1}^{l-u-1} \sum_{1, ..., u} \sum_{u+1, ..., l} (-1)^{l-\alpha-\beta+1} x^{i_1, ..., i_\beta, j_1, ..., j_\alpha}_{b_{i_1}, ..., b_{i_\beta}, a_{j_1}, ..., a_{j_\alpha}} . \qquad ... (3.8)$$

Now

$$\sum_{a_{u+1} > s_{u+1}} ... \sum_{a_l > s_l} A^{1, ..., u, u+1, ..., l}_{b_1, ..., b_u, a_{u+1}, ..., a_l}$$

$$= (-1)^{l-u} \sum_{a_{u+1} \leqslant s_u} ... \sum_{a_l \leqslant s_l} A^{1, ..., u, u+1, l}_{b_1, ..., b_u, a_{u+1}, ..., a_l} = 0, \qquad ... (3.9)$$

since $T-S$ does not contain observations with $b_1 \leqslant s_1, ... b_u \leqslant s_u, a_{u+1} \leqslant s_{u+1}, ..., a_l \leqslant s_l$.

We now put

$$(t_{j_1} - s_{j_1})...(t_{j_a} - s_{j_a})\, S^{i_1,\,...,\,i_\beta,\,j_1,\,...,\,j_a}_{b_1,\,...,\,b_\beta}$$

$$= \sum_{a_1 > s_{j_1}} \cdots \sum_{a_a > s_{j_a}} x^{i_1,\,...,\,i_\beta,\,j_1,\,...,\,j_a}_{b_1,\,...,\,b_\beta,\,a_1,\,...,\,a_a}$$

and sum (3.8) over all values $a_{u+1} > s_{u+1},\, ...,\, a_l > s_l$, divide by $(-1)^{l-u-1}$, note that

the first double sum in (3.8) is equal to $A^{1,\,...,\,u}_{b_1,\,...,\,b_u}$ and rearrange to get

$$A^{1,\,...,\,u}_{b_1,\,...,\,b_u} = \sum_{\beta=0}^{u}(-1)^{u-\beta}\sum_{a=1}^{l-u}(-1)^{a+1}\sum_{1,\,...,\,u}\sum_{u+1,\,...,l}S^{i_1,\,...,\,i_\beta,\,j_1,\,...,\,j_a}_{b_{i_1},\,...,\,b_{i_\beta}}. \quad \ldots (3.10)$$

From (3.8) we now get for $a_{u+1} > s_u,\, ...,\, a_l > s_l$

$$A^{1,\,...,\,u,\,u+1,\,...,\,l}_{b_1,\,...,\,b_u,\,a_{u+1},\,...,a_l}$$

$$= (-1)^{l-u}\, A^{1,\,...,\,u}_{b_1,\,...,\,b_u} + \sum_{\beta=0}^{u}\sum_{a=1}^{l-u}\sum_{1,\,...,\,u}\sum_{u+1,\,...,\,l}(-1)^{l-a-\beta}x^{i_1,\,...,\,i_\beta,\,j_1,\,...,\,j_a}_{b_{i_1},\,...,\,b_{i_\beta},\,a_{j_1},\,...,\,a_{j_a}}.$$

$$\ldots (3.11)$$

Equations (3.10) and (3.11) show that the solutions $A^{i_1,\,...,\,i_a}_{a_1,\,...,\,a_a}$ are completely

determined by the values $x_{a_1,\,...,\,a_n}$ and the restrictions (1.7). On the other hand,

the values of $A^{i_1,\,...,\,i_a}_{a_{i_1},\,...,\,a_{i_a}}$ also completely determine $x_{a_1,\,...,\,a_n}$. Hence the degrees of

freedom of Q_a are equal to the number of independent parameters put equal to 0 by

the assumption. If $s_1 < t_1,\, ...,\, s_l < t_l,\, s_{l+1} = t_{l+1},\, ...,\, s_n = t_n$ then there are in T

just $\prod_{i=1}^{n}(t_i-1)$ independent parameters associated with the highest order interaction

and of these $\prod_1^l s_i \prod_{l+1}^n (t_i-1)$ do not occur in $T-S$. Hence Q_a has $\prod_{l+1}^n (t_i-1)$

$\left(\prod_1^l (t_i-1) - \Pi\, s_i \right)$ degrees of freedom.

We shall give the solutions explicitly for the cases $n = 2$ and $n = 3$. In the

following b_i will denote a number $\leqslant s_i$, a_i a number $> s_i$.

$$n = 2,\, s_1 < t_1,\, s_2 < t_2;$$

$$A = S^1 + S^2 - S^{12},$$

$$A^1_{b_1} = S^{12}_{b_1} - S^2 = x^1_{b_1} - \frac{1}{t_2-s_2}\sum_{s_2+1}^{t_2} x^2_{a_2}.$$

$$A^1_{a_1} = x^1_{a_1} - A.$$

193

The case $s_2 = t_2$ can be treated as a complete factorial design with the levels of the first factor running from $s_1 + 1$ to t_1.

$$n = 3, \quad s_1 < t_1, \quad s_2 < t_2, \quad s_3 < t_3 :$$

$$A = S^1 + S^2 + S^3 - S^{12} - S^{13} - S^{23} + S^{123},$$

$$\underset{b_1}{\overset{1}{A}} = -\underset{b_1}{\overset{123}{S}} + \underset{b_1}{\overset{12}{S}} + \underset{b_1}{\overset{13}{S}} - S^2 - S^3 + S^{23},$$

$$\underset{a_1}{\overset{1}{A}} = \underset{a_1}{\overset{1}{x}} - A,$$

$$\underset{b_1 a_2}{\overset{12}{A}} = -\underset{b_1}{\overset{1}{A}} - \underset{a_2}{\overset{2}{x}} + \underset{b_1 a_2}{\overset{12}{x}},$$

$$\underset{b_1 b_2}{\overset{12}{A}} = -\underset{b_1 b_2}{\overset{123}{S}} - \underset{b_2}{\overset{23}{S}} - \underset{b_1}{\overset{13}{S}} + S^3,$$

$$\underset{a_1 a_2}{\overset{12}{A}} = \underset{a_1 a_2}{\overset{12}{x}} - \underset{a_1}{\overset{1}{x}} - \underset{a_2}{\overset{2}{x}} + A.$$

If $t_1 < s_1, t_2 < s_2, t_3 = s_3$ then in the above formulae all terms containing the index 3 must be deleted. But $\underset{b_2}{\overset{3}{A}}, \underset{a_1 b_3}{\overset{13}{A}}$ and $\underset{b_1 b_3}{\overset{13}{A}}$ can be obtained from the above formulae by permutation of indices.

For $t_2 = s_2, t_3 = s_3$ we get a complete design with the first index varying only from $s_1 + 1$ to t_1.

As can be seen the computation is not excessive and is well adapted to punch card machines.

The procedure given for testing the linear hypothesis (2.1) can be adapted also to test the following more general linear hypothesis : Let I be a natural set of interactions and $i_1, \ldots, i_u$ an interaction of I such that $I - (i_1, \ldots, i_u) = I_1$ is also a natural set. Under the assumption that all interactions outside of I are 0 in T and that all interactions of $(i_1, \ldots, i_u)$ are 0 in $T - S$, we wish to test the hypothesis that all interactions outside I and the interactions $(i_1, \ldots, i_u)$ are 0 in T. In symbols

$$A: \quad x_{a_1, \ldots, a_n} = \sum_I \mu^{i_1, \ldots, i_\alpha}_{a_{i_1}, \ldots, a_{i_\alpha}} \quad \text{in} \quad T,$$

$$x_{a_1, \ldots, a_n} = \sum_{I_1} \bar{\mu}^{i_1, \ldots, i_\alpha}_{a_{i_1}, \ldots, a_{i_\alpha}} \quad \text{in} \quad T - S. \qquad \ldots \quad (3.12)$$

$$H: \quad x_{a_1, \ldots, a_n} = \sum_{I_1} \mu^{i_1, \ldots, i_\alpha}_{a_{i_1}, \ldots, a_{i_\alpha}} \quad \text{in} \quad T.$$

Without loss of generality we may assume $i_1, ..., i_u = (1, ..., u)$. Let $\bar{I}_1, \bar{I}$ denote the interactions not in I_1, I respectively. From the identity 5.9 of Mann (1949) we get

$$Q_r = \sum_{\bar{I}_1} \frac{t_1 t_2 \cdots t_n}{t_{i_1} \cdots t_{i_\alpha}} \sum_{a_1} \cdots \sum_{a_n} \left(A^{i_1, ..., i_\alpha}_{a_1, ..., a_\alpha} \right)^2,$$

$$Q_a = \sum_{\bar{I}} \frac{t_1 t_2 \cdots t_n}{t_{i_1} \cdots t_{i_\alpha}} \sum_{a_1} \cdots \sum_{a_\alpha} \left(A^{i_1, ..., i_\alpha}_{a_1, ..., a_\alpha} \right)^2 + t_{u+1} \cdots t_n \min$$

$$\sum_{1, ..., u} \frac{t_1 \cdots t_u}{t_{i_1} \cdots t_{i_\alpha}} \sum_{a_1} \cdots \sum_{a_\alpha} \left(A^{i_1, ..., i_\alpha}_{a_1, ..., a_\alpha} - \mu^{i_1, ..., i_\alpha}_{a_1, ..., a_\alpha} \right)^2;$$

where the minimum must be computed under the assumptions A. Now again by 5.9 of Mann (1949).

$$\sum_{1, ..., u} \frac{t_1 \cdots t_u}{t_{i_1} \cdots t_{i_\alpha}} \sum_{a_1} \cdots \sum_{a_\alpha} \left(A^{i_1, ..., i_\alpha}_{a_1, ..., a_\alpha} - \mu^{i_1, ..., i_\alpha}_{a_1, ..., a_\alpha} \right)^2$$

$$= \sum_{a_1} \cdots \sum_{a_u} \left(x^{1, ..., u}_{a_1, ..., a_u} - \sum_{\alpha=1}^{u} \sum_{1, ..., u} \mu^{i_1, ..., i_\alpha}_{a_1, ..., a_\alpha} \right)^2$$

so that

$$Q_r - Q_a = t_{u+1} \cdots t_n \left[\sum_{a_1} \cdots \sum_{a_u} \left(A^{1, ..., u}_{a_1, ..., a_u} \right)^2 \right.$$

$$\left. - \min \sum'_{a_1} \cdots \sum'_{a_u} \left(x^{1, ..., u}_{a_1, ..., a_u} - \sum_{\alpha=0}^{u-1} \sum_{1, ..., a_u} \mu^{i_1, ..., i_\alpha}_{a_1, ..., a_\alpha} \right)^2 \right],$$

where Σ' is extended over those combinations $a_1, ..., a_u$ only which occur in $T-S$. In minimizing the second sum of squares we can use formulae (3.10) and (3.11) substituting u for n and $x^{1, ..., u}_{a_1, ..., a_u}$ for $x_{a_1, ..., a_n}$. It should be noted that $Q_r - Q_a = 0$ unless either $n = u$ or $t_{u+1} = s_{u+1}, ..., t_n = s_n$, since otherwise all combinations $a_1, ..., a_u$ which occur in T will also occur in $T-S$. The degrees of freedom for Q_a are obtained in the same manner as for the linear hypothesis (2.1). If in $T-S$ we have $1 \leqslant a_1 \leqslant t_1, ..., 1 \leqslant a_u \leqslant t_u$ with exception of the set $1 \leqslant a_1 \leqslant s_1, ..., 1 \leqslant a_u \leqslant s_u$ and if $s_1 < t_1, ..., s_l < t_l, s_{l+1} = t_{l+1}, ..., s_u = t_u$ then Q_a has $\prod_{l+1}^{u} (t_j - 1) \left[\prod_1^l (t_j - 1) \right.$

$\left. - \prod_1^l s_j \right]$ degrees of freedom.

4. Tests when *a priori* information is not available

The test of the hypothesis (3.12) can be carried out with only one observation, in each cell. The situation leading to the test (2.4) arises in particular in Case II. In this case we wish to test hypotheses H_i, including hypotheses suggested by the data that the means of certain linear forms whose sum of squares equals $Q_r - Q_a = Q_{S_i}$ are 0, where $Q_r - Q_a$ is computed from a set S_i. We shall be able to determine a constant h with the property that

$$P = P(Q_{S_i} \leqslant h\hat{\sigma}^2) \geqslant 1 - \alpha, \qquad \ldots \quad (4.1)$$

where P is the probability that the inequality $Q_{S_i} \leqslant h\hat{\sigma}^2$ holds simultaneously for all S_i for which H_i is true. Alternately we may say that the probability that a true hypothesis will be rejected is at most α. It should be noted that we cannot possibly make a statement about the exact probability of rejecting a true hypothesis, since this probability depends on the number of hypotheses which are true. For instance if all H_i are false the probability of rejecting a true hypothesis is 0. It should be recognized, however, that the same situation also obtains when we test a single hypothesis.

A similar problem was treated by Scheffé (1953), where the statements made concerned the mean values of individual linear forms. We shall need a generalization of Scheffé's theorem.

Theorem 2 : Let R_k be a space of normally distributed random variables of dimension k. (That is to say there exist in R_k random variables $\theta_1, \ldots, \theta_k$ such that $|\sigma_{\theta_i \; \theta_j}| \neq 0$, while for any $k+1$ random variables $\theta_1, \ldots, \theta_{k+1}$ we have $|\sigma_{\theta_i \; \theta_j}| = 0$.) Suppose $\sigma_{\theta_i \; \theta_j} = c_{\theta_i \; \theta_j} \sigma^2$ where $c_{\theta_i \; \theta_j}$ is assumed to be known and σ^2 is an unknown constant. Suppose further that $\hat{\sigma}^2$ is an estimate of σ^2 such that $v\hat{\sigma}^2$ has the χ^2 distribution with v d.f. and is independent of the variables in R_k. We shall call θ normalized if $\sigma_\theta^2 = \sigma^2$. Let $F_\alpha (k, v)$ be the upper α point of the F distribution with k and v d.f.; then, if $0 < r_1 < \ldots < r_s \leqslant k$,

$$P = P\{(\theta_1 - E(\theta_1))^2 + \ldots + (\theta_{r_i} - E(\theta_{r_i}))^2 \leqslant kF_\alpha(k, v)\hat{\sigma}^2\} = 1 - \alpha, \qquad \ldots \quad (4.2)$$

where P is the probability that the inequality in parenthesis is true for all r_i and all choices of r_i independent normalized random vectors of R_k.

Proof: Let $\eta_1, \ldots, \eta_k$ be a basis of R_k such that $\sigma_{\eta_i \; \eta_j} = \text{diag. } \sigma^2$. If $\theta = a_1 \eta_1 + \ldots + a_k \eta_k$ then

$$\theta - E(\theta) = \Sigma \; a_i(\eta_i - E(\eta_i)),$$

$$\sigma_\theta^2 = \Sigma \; a_i^2 \; \sigma^2.$$

If θ is normalized then $\Sigma \; a_i^2 = 1$. Hence our theorem will follow from the following lemma.

Lemma 2 : *Let $y_1, ..., y_k$ be k real numbers. Consider for every r all sets of r normalized orthogonal forms $l_i = \sum_j d_{ij} y_j$, $i = 1, ..., r$, $\sum_j d_{ij} d_{lj} = \delta_{il}$. If $\sum y_i^2 \leqslant a^2$ then $l_1^2 + ... + l_r^2 \leqslant a^2$ for all choices r and all choices $l_1, ..., l_r$. If $l_1^2 + ... + l_r^2 \leqslant a^2$ for one fixed r and all choices $l_1, ..., l_r$, then $\sum y_i^2 \leqslant a^2$.*

Proof : Let $\sum y_i^2 \leqslant a^2$. We can augment the $r \times k$ matrix (d_{ij}) to an orthogonal $k \times k$ matrix (d_{ij}). We then have $\sum_1^k y_i^2 = \sum_1^k l_i^2$. Hence $\sum_1^k y_i^2 \leqslant a^2$ implies $\sum_1^r l_i^2 \leqslant a^2$.

Let $\sum_1^r l_i^2 \leqslant a^2$ for all choices $l_1, ..., l_r$ (r fixed). We determine an orthogonal $(k-r) \times k$ matrix (a_{ij}) such that $\sum_j a_{ij} y_j = 0$. We augment this matrix to an orthogonal $k \times k$ matrix. Then $\sum_1^k y_i^2 = \sum_{k-r+1}^k l_i^2 \leqslant a^2$.

The theorem now follows immediately since by Lemma 2

$$P = P((\theta_1 - E(\theta_1))^2 + ... + (\theta_{r_i} - E(\theta_{r_i}))^2 \leqslant k F_a(k, v) \hat{\sigma}^2)$$

$$= P((\eta_1 - E(\eta_1))^2 + ... + (\eta_k - E(\eta_k))^2 \leqslant k F_a(k, v) \hat{\sigma}^2) = 1 - \alpha.$$

By p. 142 of Mann (1949), we can find an orthogonal system $l^{(\gamma)}_{i_1, ..., i_\alpha}$, $\gamma = 1, ..., \prod_1^\alpha (t_{i_\beta} - 1)$, $\alpha = 0, ..., n$ and $i_1, ..., i_\alpha$ runs through all choices $i_1 < ... < i_\alpha$ out of $1, ..., n$, where $l^{(\gamma)}_{i_1, ..., i_\alpha}$ is a component of the interaction between the factors $i_1, ..., i_\sigma$ (for definition see Mann (1949, p. 140). Every linear form can then be represented in the form

$$l = \sum_{\alpha=0}^n \sum_{1, ..., n} \sum_\gamma a^{(\gamma)}_{i_1, ..., i_\alpha} l^{(\gamma)}_{i_2, ..., i_\alpha} \qquad ... \quad (4.3)$$

We modify the definition of interaction component so that 0 is a component of every interaction then

Theorem 3 : *Let l be a linear form of the observations such that $E(l) = 0$ if $\mu^{i_1, ..., i_\alpha}_{a_1, ..., a_\alpha} = 0$ for a set I of interactions $i_1, ..., i_\alpha$ and all possible values $a_{i_1}, ..., a_{i_\alpha}$. Then l is a linear combination of components of interactions from the set I.*

Proof : It is easy to see that the expectation of a component of the interaction between $i_1, ..., i_\alpha$ is a linear form in the $\mu^{i_1, ..., i_\alpha}_{a_1, ..., a_\alpha}$. Hence by assumption

$$E(l) = \sum_{\bar{I}} \sum_\gamma a^{(\gamma)}_{i_1, ..., i_\alpha} E\left(l^{(\gamma)}_{i_1, ..., i_\alpha} \right) = 0.$$

Since $E\left(l_{i_1, \ldots, i_\alpha}^{(\gamma)}\right)$ are independent linear forms in $\mu_{a_1, \ldots, a_\alpha}^{i_1, \ldots, i_\alpha}$ we can for $i_1, \ldots, i_\alpha \epsilon \bar{I}$

give them any arbitrary values, whence $a_{i_1, \ldots, i_\alpha}^{(\gamma)} = 0$ for $(i_1, \ldots, i_\alpha) \epsilon \bar{I}$ and all values

of γ, which is just the assertion of Theorem 3.

Theorem 4: *If* $E(x_a) = \sum\limits_{i=1}^{s} g_{ai}\beta_i$, $\alpha = 1, \ldots, N$ *and* $\hat{\beta}_i$ *are (not necessarily*

unique) least square estimates of β_i *then* $E\left(\sum g_{ai}\hat{\beta}_i\right) = E(x_a)$.

   Proof : Let

$$G = (g_{ai}), = x \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix}, \beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_s \end{bmatrix}.$$

The assumption then reads $E(x) = G\beta$. The least square equations read $G'x = G'G\hat{\beta}$ and $\hat{\beta} = Px$ where the matrix P satisfies the equation $G'GP = G'$. Now GP is symmetric [Mann, (1960), Theorem 1] and hence $GPG = P'G'G \equiv G$. Therefore

$$E(x - G\hat{\beta}) = E(x - GPx) = E(x) - GPG\beta = E(x) - G\beta = 0.$$

Theorem 5 : *Let* A *be the region* $1 \leqslant a_1 \leqslant t_1, \ldots, 1 \leqslant a_n \leqslant t_n$ *and let* B

be any subregion of A. *Let* I *be a natural set of interactions and let* $A_{a_1, \ldots, a_\alpha}^{i_1, \ldots, i_\alpha}$

$(i_1, \ldots, i_\alpha) \epsilon I$ *be the least square estimates obtained by minimizing* $\sum\limits_{B}\Big(x_{a_1, \ldots, a_n}$

$-\sum\limits_{I} \mu_{a_{i_1}, \ldots, a_{i_\alpha}}^{i_1, \ldots, i_\alpha}\Big)^2$ *and let* $X_{a_1, \ldots, a_n}$ *be the corresponding regression values. Then*

$$x_{a_1, \ldots, a_n} - X_{a_1, \ldots, a_n}$$

is a linear combination of interaction components of $\bar{I}$.

   Proof : Since I is a natural set

$$E(x_{a_1, \ldots, a_n}) = \sum\limits_{I} \mu_{a_1, \ldots, a_\alpha}^{i_1, \ldots, i_\alpha} \text{ in } A \qquad \ldots \text{(4.4)}$$

implies the same representation in B. Hence $E(X_{a_1, \ldots, a_n}) = E(x_{a_1, \ldots, a_n})$ if (4.4) holds in A. Theorem 5 now follows from Theorem 3.

   Now consider a situation where an independent estimate $\hat{\sigma}^2$ of σ^2 is available and we compute $Q_{S_i} = Q_r - Q_a$ for various regions S_i including some suggested by the data, where Q_r and Q_a are computed under the linear hypothesis (3.12) with S replaced by S_i. The computation shows that Q_{S_i} is a sum of squares of linear forms in the $x_{a_1, \ldots, a_u}^{1, \ldots, u}$ which themselves are linear forms in $A_{a_1, \ldots, a_\alpha}^{i_1, \ldots, i_\alpha}$, $i_\alpha \leqslant u$. It follows

thus from Theorem 5 that $Q_{S_i} = \sum\limits_{j=1}^{u_i} l_{ij}^2$ where the l_{ij} are components of the interaction $(1, \ldots, u)$. The space of these interactions has the dimension

$$k = \prod_1^u (t_j - 1).$$

Now let H_i be the hypothesis $E(l_{ij}) = 0$, $j = 1, \ldots, u_i$. Then by Theorem 2

$$P = P(Q_{S_i} < k F_a(k, v) \hat{\sigma}^2) \geqslant 1 - \alpha, \qquad \ldots \quad (4.5)$$

where P is the probability that the inequality in parenthesis holds for all values i for which H_i is true.

It should be noted that Q_{S_i} for every S_i is a summand of the $(1, \ldots, u)$ interaction sum of squares of T. Hence if H_i is rejected for any i then the $(1, \ldots, u)$ interaction sum of squares in T will itself be significant on the level α and (4.5) will serve only to locate regions S_i from which a significant contribution to the $(1, \ldots, u)$ interaction sum of squares in T arises. The situation is quite different in case I, where we test only one region S. In this case the value of k in (4.5) is the number of degrees of freedom of Q_S and it may very well happen that the hypothesis H of (3.12) will be rejected even if the $(1, \ldots, u)$ interaction sum of squares in T is not significant on the level α.

One is tempted also to compare $Q_r - Q_a$ and Q_a of the hypothesis (3.12), even if the assumption is not known to be true. If the statistic F of (2.2) is large this would be taken to mean that the contribution of the region S to the interaction sum of squares is large compared to that of $T - S$. However, it should be kept in mind that F is then the ratio of two non-central χ^2 and thus does not have the F distribution. The value of F in this case may be very useful as a basis for conjectures and as a lead for further research, but it cannot be used for an exact test.

5. POWER OF THE TEST

We shall now show that the test (2.4) is at least as powerful as the test for highest order interactions in T if the interactions in $T - S$ are 0.

We consider regions W in the n-dimensional Euclidean space R_n and two distribution-functions F_1, F_2 in R_n. The integral $\int\limits_W dF_1 = P_1(W)$ is called the size of W. The integral $\int\limits_W dF_2 = P_2(W)$ will be called the power of W.

A family L of regions in R_n will be called an additive family if sums, intersections and differences of L are again in L.

In the following definitions and theorems all regions considered will be regions of an additive family L. To avoid cumbersome language we shall however just speak of regions. A most powerful region W will mean a region of the family L such that $P_2(W) \geqslant P_2(W')$ for all $W' \epsilon L$ for which $P_1(W) = P_1(W')$. We shall also assume that our regions satisfy the following conditions.

(i) *If W is any region of size α and if $\beta < \alpha$ then W has a subregion of size β.*

Condition (i) obviously implies

(i') *If W is any region of size α and if $\alpha = \overset{n}{\underset{i=1}{\Sigma}} \beta_i$ then $W = \Sigma W_i$, where W_i has size β_i.*

Lemma 3 : *Let W be a most powerful region of size α and W_1 any subregion of W of size $\beta \leqslant \alpha$. Let K be any region of size β and $K \cap W$ empty. Then $P_2(K) \leqslant P_2(W_1)$.*

Otherwise the region $W - W_1 + K$ would have size α and higher power than W.

Lemma 4 : *Let W be a most powerful region of size α and W^* any subregion of W of size $\beta \leqslant \alpha$. Let K be any region of size $k\beta$ such that $K \cap W$ is empty, then*

$$kP_2(W^*) \geqslant P_2(K). \qquad \ldots (5.1)$$

Proof : Choose $\epsilon > 0$ and arbitrary. Put $n = \left[\dfrac{1}{\epsilon}\right] + 1$. Divide K into n regions of size $\dfrac{k\beta}{n}$. In at least one of these regions, say in D_1, we must have $P_2(D_1) \leqslant \epsilon$.

Hence if $0 \leqslant \delta \leqslant \dfrac{k\beta}{n}$ there is a region $D \subset K$ such that $P_1(D) = \delta$, $P_2(D) \leqslant \epsilon$.

Now let $\beta = m\delta$, m integral $k\beta = km\delta = [km]\delta + \delta_1$ where $\delta_1 < \delta$. Let $W^* = \overset{m}{\underset{1}{\Sigma}} W_i$ where W_i has size δ. Let $K = \Sigma K_i + D$, where $P_1(K_i) = \delta$, $P_1(D) = \delta_1$, $P_2(D) \leqslant \epsilon$. Let $p = \min P_2(W_i)$. By Lemma 3 we have $p \geqslant P_2(K_i)$, whence

$$P_2(K) \leqslant [km]p + \epsilon \leqslant kP_2(W^*) + \epsilon.$$

Since ϵ is arbitrary (5.1) must hold.

Theorem 6 : *Let W be a most powerful region of size α. Let $W_1, \ldots, W_t$ be t regions $P_1(W_i) = \alpha_i$ and let $p_1, \ldots, p_t$ be t non-negative numbers such that*

$$\Sigma p_i \alpha_i = \alpha, \quad \Sigma p_i = 1.$$

Then

$$\underset{i}{\Sigma} \, p_i P_2(W_i) \leqslant P_2(W).$$

Proof : Suppose first that $W \not\subset W_1$, $W_1 \not\subset W$. Let $P_1(W - W \cap W_1) = \beta_1$, $P_1(W_1 - W \cap W_1) = \beta_2$. Suppose $\beta_1 \geqslant \beta_2$. Then $P_2(W_1)$ is not decreased if W_1 is replaced by the region $W_1 \cap W + W^*$ where W^* is a subregion of $(W - W \cap W_1)$ of size β_2. If $\beta_1 < \beta_2$ choose $W^* \subset W_1 - W \cap W_1$ of size β_1 and replace it by $W - W_1 \cap W$.

Without loss of generality we may, therefore, assume that either $W_i \supset W$ or $W_i \subset W$.

Now suppose $W_1 \supset W$, $W_2 \subset W$. Let $P_1(W_1 - W) = \beta_1$, $P_1(W - W_2) = \beta_2$. If $p_1\beta_1 \geqslant p_2\beta_2$ replace W_2 by W and subtract from $W_1 - W$ a region W^* of size $\dfrac{p_2}{p_1}\beta_2$. Let the two new regions be W'_1, W'_2. Then

$$p_1 P_1(W'_1) + p_2 P_1(W'_2) = p_1 P_1(W_1) + p_2 P_1(W_2)$$

and (Lemma 4)

$$p_1 P_2(W_1') + p_2 P_2(W_2') = p_1 P_2(W_1) + p_2 P_2(W_2) - p_1 P_2(W^*) + p_2 P_2(W - W_2)$$
$$\geqslant p_1 P_2(W_1) + p_2 P_2(W_2).$$

If $p_1 \beta_1 < p_2 \beta_2$ replace W_1 by W and add a region $W^* \subset W - W_2$ of size $p_1 \beta_1 / p_2$ to W_2.

We can continue this process until either $W_i \supset W$ for all W_i or $W \subset W_i$ for all W_i, but then the equation $\Sigma p_i P_1(W_i) = \alpha$ shows that all W_i are of size α and we can replace them by W. This proves Theorem 6.

Corollary to Theorem 6 : *Let $\{W_i\}$ be an infinite sequence of regions and, $\{p_i\}$ a sequence of non-negative numbers such that $\Sigma p_i P_1(W_i) = P_1(W)$ and $\Sigma p_i = 1$, where W is a most powerful region. Then*

$$\Sigma p_i P_2(W_i) \leqslant P_2(W).$$

Proof: Find N so that $\overset{\bullet}{\underset{N+1}{\Sigma}} p_i \leqslant \epsilon$. Then

$$\Sigma p_i P_2(W_i) \leqslant \overset{N}{\underset{1}{\Sigma}} p_i P_2(W_i) + \epsilon \leqslant P_2(W) + \epsilon.$$

Theorem 7 : *Let S be a space and Q a probability measure defined over S. For every $z \epsilon S$ let $W(z)$ be a region in R_n. Let further*

$$\int P_1(W(z)) dQ = \alpha$$

and assume that $\int P_2(W(z)) dQ$ exists. If W is a most powerful region of size α, then

$$\int P_2(W(z)) dQ \leqslant P_2(W). \qquad \dots (5.2)$$

For every Q measurable set S_i let $Q(S_i)$ be the Q measure of S_i. For every $\epsilon_1 > 0, \epsilon > 0$ we can find a covering $\{S_i\}$ of S such that

$$\int P_2(W(z)) dQ = \Sigma P_2(W(\xi_i)) Q(S_i) - \eta_1$$

$$0 \leqslant \eta_1 \leqslant \epsilon_1, \qquad \xi_i \epsilon S_i$$

and

$$\alpha = \int P_1(W(z)) dQ = \Sigma P_1(W(\xi_i)) Q(S_i) + \eta, \qquad |\eta| < \epsilon\alpha.$$

If η is positive apply the Corollary to Theorem 6 to a most powerful subregion of W of size $\alpha - \eta$. If η is negative, choose a most powerful region W^* of size $\alpha + \eta$ containing W. Then

$$\int P_2(Wz) dQ \leqslant \Sigma P_2(W)(\xi_i) Q(S_i) \leqslant P_2(W^*) = P_2(W) + P_2(W^* - W) \leqslant P_2(W) + \epsilon.$$

This proves (5.2).

We now return to the test (2.4). The highest interaction sum of squares in T is our Q_r and we have

$$Q_r = Q_r - Q_a + Q_a = \Sigma m_i^2 + \Sigma l_i^2, \qquad \dots (5.3)$$

where the l_i are normalized orthogonal forms whose means are 0 by assumption and the hypothesis H states that $E(m_i) = 0$. In the test (2.4) we form $F = c \dfrac{\Sigma m_i^2}{v\hat{\sigma}^2} = c\dfrac{\chi_1^2}{\chi_3^2}$. In testing the interaction sum of squares in T we form

$F_1 = c_1 \dfrac{\chi_1^2 + \chi_2^2}{\chi_3^2}$, where $\chi_2^2 = \dfrac{1}{\sigma^2} \Sigma l_i^2$. We shall show that the region $F \geqslant F_0$ or $\chi_1^2 \geqslant k\chi_3^2$ is at least as powerful as the region $F_1 \geqslant F_{10}$ with respect to all alternatives $\Sigma E(m_i)^2 > 0$.

The region $F_1 \geqslant F_{10}$ is equivalent to

$$\chi_1^2 > k_1\chi_3^2 - \chi_2^2, \qquad \qquad \dots \quad (5.4)$$

where $$P(\chi_1^2 \geqslant k\chi_3^2) = P(\chi_1^2 \geqslant k_1\chi_3^2 - \chi_2^2)$$

provided $E(m_i) = 0$. We shall use a result of P. L. Hsu (1941) that the critical region $F \geqslant F_0$ maximizes the power among all regions W of equal size, whose power depends only on $\lambda = \Sigma(E(m_i))^2$. Consider all regions (5.4) for varying values of χ_2^2 as regions in the space χ_1^2, χ_3^2 and consider all regions obtainable from them by the operations of addition, intersection and subtraction. The power function of any region of this additive family L of regions depends only on λ (see f.i. Mann (1949) pp. 65, 66). Hence if $P(F \geqslant F_0) = \alpha$ then the region $F \geqslant F_0$ is a most powerful region of size α of the family L. It is not difficult to check that the condition (i) of Theorem 7 is satisfied by L. Now in Theorem 7 let F_1 be the d.f. of χ_1^2 if $E(m_i) = 0$ and F_2 its distribution if $\Sigma E(m_i)^2 = \lambda \neq 0$. Let $W(\chi_2^2)$ be the region $\chi_1^2 \geqslant k_1\chi_3^2 - \chi_2^2$ and let Q be the cumulative function of χ_2^2. Then Theorem 7 shows immediately that the region $F \geqslant F_0$ is at least as powerful as a region (5.4) or the equivalent region $F_1 \geqslant k_1$ of the same size. Actual calculation in special cases, using the tables in Tang (1938) or Mann (1949), shows that there is actually an appreciable gain in power.

REFERENCES

Hsu, P. L. (1941): Analysis of variance from the power-function standpoint. *Biometrika*, **32**, 62–69.

Mann, H. B. (1949): *Analysis and Design of Experiments*. Dover Publications, Inc.

———— (1960): The algebra of a linear hypothesis. *Ann. Math. Stat.*, **31**, 1–15.

Scheffé, Henry (1953): A method for judging all contrasts in the analysis of variance. *Biometrika*, **40**, 87–104.

Tang, P. C. (1938): The power function of the analysis of variance tests with tables and illustrations of their use. *Statistical Research Memoirs*, **2**, 126–149 and tables.

Paper received : November, 1960.

LISTING OF BIB DESIGNS FROM $r = 16$ TO 20

By D. A. SPROTT

University of Waterloo

v	b	r	k	λ		v	b	r	k	λ	
241	241	16	16	1		37	222	18	3	1	T_2
121	121	16	16	2	*	91	273	18	6	1	
81	81	16	16	3		353	353	19	19	1	*
61	61	16	16	4		172	172	19	19	2	*
49	49	16	16	5		115	115	19	19	3	
41	41	16	16	6		58	58	19	19	6	*
31	31	16	16	8	C	39	39	19	19	9	
225	240	16	15	1		324	342	19	18	1	
65	80	16	13	3		153	171	19	17	2	
177	236	16	12	1		96	114	19	16	3	
45	60	16	12	4		39	57	19	13	6	
33	48	16	11	5		20	38	19	10	9	4
145	232	16	10	1		153	323	19	9	1	
25	40	16	10	6		96	304	19	6	1	
113	226	16	8	1		20	76	19	5	4	
29	58	16	8	4	C	20	95	19	4	3	γ_1
17	34	16	8	7	A	39	247	19	3	1	T_1
81	216	16	6	1		381	381	20	20	1	$P(2,19):1$
21	56	16	6	4	**	191	191	20	20	2	
65	208	16	5	1	G	96	96	20	20	4	
49	196	16	4	1		77	77	20	20	5	*
17	68	16	4	3	X_2	39	39	20	20	10	
33	178	16	3	1	T_1	361	380	20	19	1	$E(2,19):1$
273	273	17	17	1	$P(2,16):1$	171	190	20	18	2	
137	137	17	17	2*							
69	69	17	17	4		76	95	20	16	4	
35	35	17	17	8		141	188	20	15	2	
256	272	17	16	1	$E(2,16):1$	57	76	20	15	5	
						36	48	20	15	8	
120	136	17	15	2							
52	68	17	13	4							
18	34	17	9	8		21	28	20	15	14	*
120	255	17	8	1		111	185	20	12	2	
						45	75	20	12	5	
35	85	17	7	3	.	21	35	20	12	11	
18	51	17	6	5							
35	119	17	5	2	β_2						
52	221	17	4	1	F_3	181	362	20	10	1	
						61	122	20	10	3	
18	102	17	3	2	l_1	46	92	20	10	4	
307	307	18	18	1	$P(2,17):1$	37	74	20	10	5	C
154	154	18	18	2							
103	103	18	18	3	*						
						21	42	20	10	9	B_1
52	52	18	18	6	*	13	26	20	10	15	B_3
35	35	18	18	9		36	90	20	8	4	
289	306	18	17	1	$E(2,17):1$	51	170	20	6	2	
136	153	18	16	2							

* Non existent.

Series α_1, α_2, β_2, γ_1 from Bose (1942).

Series C, A, O_2, B_1, B_3, A_1 from Sprott (1954).

** Derived from the Steiner system $S(5,8,24)$, Sprott (1955).

Series G_2, T_1, F_3, E_1, T_2 from Bose (1939).

Series l_1, 1, 4 from Sprott (1956).

LISTING OF BIB DESIGNS FROM $r = 16$ TO 20—(continued)

v	b	r	k	λ		v	b	r	k	λ	
85	102	18	15	3		21	70	20	6	5	
100	150	18	12	2		81	324	20	5	1	
34	51	18	12	6		17	68	20	5	5	1
55	99	18	10	3		61	305	20	4	1	
145	290	18	9	1							
49	98	18	9	3		31	155	20	4	2	
17	34	18	9	9	$1, A_1$	21	105	20	4	3	
19	57	18	6	5	A	11	55	20	4	6	

REFERENCES

Bose, R. C. (1939) :   On the construction of balanced incomplete block designs. *Ann. Eugen.*, **10**, 353-399.

———— (1942) :   On some new series of balanced incomplete block designs. *Bull. Cal. Math. Soc.*, **34**, 17-31.

Rao, C. R. (1961) :   A study of BIB designs with replications 11 to 15.   *Sankhyā*, **23**, 117–127.

Sprott, D. A. (1954) :   A note on balanced incomplete block designs. *Can. J. Math.*, **6**, 341-346.

———— (1956) :   Some series of balanced incomplete block designs. *Sankhyā*, **17**, 185-192.

———— (1955) :   Balanced incomplete block designs and tactical configuration. *Ann. Math. Stat.*, **26**, 752-758.

Paper received : November, 1961.

CORRIGENDA

The Transformation of a Distribution Under Selection : By D. J. Finney, *Sankhyā*, Series A, **23**, 309-324

<div>

p. 321, line 11 from below   For $(\gamma_{14}+4\rho\gamma_{33})$ read $(\gamma_{14}+4\rho\gamma_{23})$

p. 321, line 8 from below   For $6\rho\gamma_{31}\gamma_{04}$ read $6\rho\gamma_{31}\gamma_{01}$

p. 322, line 2 from below   For $4\rho^2\gamma_{03}^2$ read $6\rho^2\gamma_{03}^2$

p. 323, line 11   For $(T^6-11T^4-25T^2-7)$ read $(T^6-11T^4+25T^2-7)$

p. 323, line 12   For $312T^2$ read $314T^2$

</div>

On a Problem of Bartholomew in Life Testing : By P. S. Swamy and S. A. D. C. Doss, *Sankhyā*, Series A, **23**, 225-230.

(1)   Equation (2.4) :   Should read as

$$k_j = \frac{pe^{-T_j/\theta_1}}{pe^{-T_j/\theta_1}+qe^{-T_j/\theta_2}} = \frac{pQ_{1j}}{Q_j}.$$

(2)   Equation (2.9) :   The last part should read as $l \neq j,\ l = j$ and not $1 \neq j,\ 1 = j$

(3)   Page 227, line 7 from below :   Should read δ_{jl} and not δ_{ll}.

(4)   Page 228, last expression :   Should read as

$$\hat{\theta} \cdot = \frac{1}{r}\left\{ \sum_{j=1}^{n} a_j t_j + (n-r)T \right\}.$$

(5)   Page 229, first expression :   Should read as

$$\hat{\theta} = \frac{1}{r}\left\{ \sum_{j=1}^{n} a_j t_j + (n-r)t_r \right\}.$$

SANKHYĀ: THE INDIAN JOURNAL OF STATISTICS

I, Prafulla Chandra Mahalanobis, hereby declare that the particulars given above are true to the best of my knowledge and belief.

15 May, 1962   *Sd/ Prafulla Chandra Mahalanobis*

BIOMETRIKA

CONTENTS

BIOMETRIKA

CONTENTS

Miscellanea : Contributions by G. G. Alway ; H. E. Daniels ; W. J. Ewens and P. D. Finch ; J. Gani ; Gerald J. Glasser ; E. J. Hannan ; Shirley E. Hitchcock ; Joseph B. Keller ; A. M. Kshirsagar ; J. L. Mott ; R. S. Pinkham ; G. P. Steck and D. B. Owen ; P. S. Swamy ; J. S. Williams ; J. Wise.

Reviews
Other Books Received
Corrigenda

Obtainable through any bookseller or direct from the

ROYAL STATISTICAL SOCIETY

21 Bentinck Street, London, W. 1.

THE ANNALS OF MATHEMATICAL STATISTICS

(Founded by H. C. Carver)

The Official Journal of the Institute of Mathematical Statistics

CONTENTS OF VOLUME 33, NO. 3, SEPTEMBER, 1962

ECONOMETRICA

CONTENTS

Ronald A Fisher

IN MEMORIAM

SIR RONALD AYLMER FISHER

(1890—1962)

Sir Ronald Fisher, F.R.S. passed away on 29 July 1962 at Adelaide, Australia after an operation. A cable communicating the news was placed before the meeting of the Council of the Indian Statistical Institute held on 31 July 1962 at New Delhi, by the President Dr. C. D. Deshmukh. Professor P. C. Mahalanobis spoke briefly, about Sir Ronald Fisher's great contributions to the advancement of science by developing statistics as a new technology which is finding increasing applications in all natural and social sciences. He laid the foundations of modern statistical theory and devised refined tools for applications. After retirement from the Cambridge University he was working as a Consultant to the Council of Scientific and Industrial Research of Australia at Adelaide. He had been in personal touch with the statistical activities in Calcutta from the early twenties and had maintained close personal contact with the workers and activities of the Indian Statistical Institute. He visited the Institute on eight occasions; the last two visits being from December 1960 to February 1961 and so recently as from December 1961 to February 1962. He was the most outstanding statistician of the world and a life-long well-wisher of the Institute.

The following resolution was passed by the Council, all members standing.

"The Council places on record the sense of profound sorrow with which it has heard of the passing away of Sir Ronald Aylmer Fisher on 29 July 1962 at Adelaide, Australia. He ushered in a new epoch in the history of statistical science and was the leader of a movement which made statistics a new technology of the present age. He established contacts with the statistical workers in Calcutta in the early twenties; served as a member of the Examination Committee of the Indian Statistical Institute in 1936; served as the Chairman of the Review Committee of the National Sample

Survey in 1956-57; and gave continuing support to the Institute at both national and international levels. He visited the Institute for the first time in 1937, and came to the Institute on seven subsequent occasions, the last visit being from December 1961 to February 1962. He was elected an Honorary Fellow of the Indian Statistical Institute in 1937, and was awarded the honorary degree of D.Sc., at the first Convocation of the Institute in February 1962 on which occasion he delivered the Convocation Address. Through his several visits to the Institute, personal contacts with its workers, scientific contributions to *Sankhyā*: the Indian Journal of Statistics, visit to other scientific centres and advisory work in India, he helped in a most significant way in the development of the integrated research, training and project programmes of the Indian Statistical Institute and in its emergence as a higher technological institution of a new type, and also generally in the advancement of statistics in India."

A condolence meeting was held on 30 July 1962 at 12 noon at the Indian Statistical Institute, 203 Barrackpore Trunk Road, Calcutta. The meeting was attended by all students and workers of the Institute. Shri S. Basu, the Joint Director and Dr. C. R. Rao, Head of the Research and Training School, spoke at the meeting.

Dr. C. R. Rao mentioned that with the passing away of Sir Ronald Fisher, the world has lost an outstanding scientist. Sir Ronald's contributions were not confined to methodological aspects of statistics which he considerably enlarged and refined, but covered applied fields such as genetics and biometry and, in general, influenced scientific thought of the present century. He laid the foundations of modern statistical theory by the introduction of small sample distributions, the theory of efficient and sufficient estimation, the likelihood principle and fiducial inference. He introduced the entirely new discipline of statistical design of experiments and analysis of variance by which scientific data are collected in an efficient way and analysed to yield valid inferences. Sir Ronald took a special interest in the teaching and research activities of the Institute and, during each of his several visits, he conducted seminars on current problems in statistics and left new ideas and new problems for the staff of the Institute to work on. During his last visit to the Institute, four months ago, he was actively working on some examples to illustrate in detail the new concepts he had introduced and to throw new light on the theory of statistical inference. A great man of science has gone, and no one knows whether the void would be filled again.

The Institute was closed for the day as a mark of respect to Sir Ronald Fisher.

DENSITY IN THE LIGHT OF PROBABILITY THEORY-II*

By E. M. PAUL

Indian Statistical Institute

SUMMARY. Let $\{X_n\}$ be a sequence of abstract spaces, each X_n consisting of the points $0, 1, 2, \ldots$. At the point r in X_n, we place probability $1/q_n^r(1-1/q_n)$, q_n being the n-th prime number. Let X be the product space $X_1 X_2 \ldots$ and let P be the product measure.

Let J be a sequence $\{j_m\}$ of positive integers. Let S be any set of positive integers. $M_J^U(S)$ is the set of vectors $(x_1, x_2, \ldots) \in X$ such that $2^{x_1} \ldots q_n^{x_n} \in S$ for infinitely many $n \in J$. $M_J^L(S)$ is the set of vectors $(x_1, x_2, \ldots) \in X$ such that $2^{x_1} \ldots q_n^{x_n} \in S$ for all sufficiently large $n \in J$. We prove that $P[M_J^L(S)] \leqslant \delta_L(S) \leqslant \delta^U(S) \leqslant P[M_J^U(S)]$ for all sets S if and only if $\frac{\log j_{m+1}}{\log j_m}$ is bounded as $m \to \infty$. δ_L and δ^U stand for lower and upper logarithmic densities, respectively.

Let f be a finite function defined on the set of positive integers. Suppose for a J satisfying the condition above, $\lim_{m \to \infty} f\left(2^{x_1} \ldots q_{j_m}^{x_{j_m}}\right) = g(x)$ exists with probability 1. Then f has a distribution and this is the same as that of $g(x)$; we employ logarithmic density.

GENERALIZATION OF THE MAGNIFICATION THEOREM

We now generalize the magnification theorem in the case of the special example discussed in the previous paper (Paul, 1962). Let J be a class of positive integers. Let S be an arbitrary set of positive integers. We define the *upper J-magnification* of S, $M_J^U(S)$, to be the set of vectors $(x_1, x_2, \ldots)$ such that $\left(2^{x_1} 3^{x_2} \ldots q_n^{x_n}\right) \in S$ for infinitely many values of $n \in J$. The *lower J-magnification* of S, $M_J^L(S)$, is defined to be the set of vectors $(x_1, x_2, \ldots)$ such that $\left(2^{x_1} 3^{x_2} \ldots q_n^{x_n}\right) \in S$ for all sufficiently large values of n in J. Obviously, $M^L(S) \leqslant M_J^L(S) \leqslant M_J^U(S) \leqslant M^U(S)$. This raises the question of obtaining sharper estimates for lower and upper logarithmic densities.

Let J consist of $j_1, j_2, \ldots$, in ascending order. We shall prove the following theorem.

Theorem : $\quad P[M_J^L(S)] \leqslant \delta^L(S) \leqslant \delta^U(S) \leqslant P[M_J^U(S)]$ *for all sets S if and only if* $\left(\frac{\log j_{n+1}}{\log j_n}\right)$ *remains bounded as* $n \to \infty$.

The proof of the 'if' part is similar to the proof given by the author (Paul, 1962). Let us call the space $X_1 X_2 \ldots X_{j_1}$ by the name Y_1 and $X_{(j_1+1)} X_{(j_1+2)} \ldots X_{j_2}$ by the name $Y_2 \ldots$ In each space X_n, let us introduce the

measure described earlier by the author (Paul, 1962) and in each space Y_m let us introduce the product measure. X may be looked upon as the space $Y_1 \, Y_2 \, Y_3 \dots$. Instead of the spaces $X_1, X_2, \dots$ (Paul, 1962; Section 2) we now have $Y_1, Y_2, \dots$. We treat the point $(0, 0, \dots, 0)$ of Y_n as the element 0 of X_n. Let $(x_1, x_2, \dots, x_n, 0, 0, \dots) \epsilon I \subset X$. We associate with it the number $q_1^{x_1} \dots q_n^{x_n}$. If $\sigma \subset I$, we define $\delta^U(\sigma)$ to be the upper logarithmic density of the corresponding set of positive integers. The space $Y_1 \, Y_2 \dots$ and δ satisfy Postulates (A) to (F) of Section 2 and condition G of Section 3 of the previous paper (Paul, 1962). The proof that condition H also holds is similar to the proof given in Section 6 of the previous paper (Paul, 1962) but requires a little explanation. Let B be a right-complete set in $I \subset Y_1 \, Y_2 \dots$. Let $(x_1, \dots, x_m, 0, 0, \dots)$ be a basic vector of B and let $x_m > 0$. Let $j_n < m \leqslant j_{(n+1)}$. Let

$$f_n(s) = \frac{(1-1/q_1^s)(1-1/q_2^s)\dots(1-1/q_{j(n+1)}^s)}{\left(2^{x_1} \dots q_m^{x_m}\right)^s}$$

We are interested in proving that $\underset{n}{\Sigma} f_n(s)$, over all basic vectors, is continuous on $[1, 2]$. Since m may be $< j_{(n+1)}$, our previous argument does not go through directly. So we introduce

$$\phi_n(s) = \frac{(1-1/q_1^s)\dots(1-1/q_m^s)}{\left(2^{x_1} \dots q_m^{x_m}\right)^s}.$$

Then

$$\frac{f_n(s)}{\phi_n(s)} > \left(1-\frac{1}{q_{j_n}}\right)\dots\left(1-\frac{1}{q_{j_{n+1}}}\right) > \frac{\log q_{j_n}}{2 \log q_{j_{(n+1)}}},$$

by Merten's theorem, $\geqslant \alpha > 0$, by hypothesis on J.

We now apply the argument given in the previous paper (Paul, 1962) and prove that $\underset{n}{\Sigma} \phi_n(s)$ is continuous on $[1, 2]$. Continuity of $\underset{n}{\Sigma} f_n(s)$ on $[1, 2]$ follows immediately, and the 'if' part is proved.

Before proving the 'only if' part, we give an example of a J and a right-complete set in $I \subset Y_1 \, Y_2 \dots$ such that condition H (Paul, 1962) is violated. Of course, in this case $\left(\frac{\log j_{n+1}}{\log j_n}\right)$ will be unbounded. Let us take a fixed number < 1, say $\frac{3}{4}$. Also, let us take the sequence $\frac{9}{10}, \frac{10}{11}, \frac{11}{12}, \dots \to 1$.

Let $j_1 = 2$, so that the first block of primes is 2, 3. Let us declare $(0, 1, 0, 0, \dots)$ and $(2, 1, 0, 0, \dots)$ as basic vectors. The cylinder sets whose bases are the points $(0, 1)$ and $(2, 1)$ carry probability

$$\beta_1 = \frac{(1-\frac{1}{2})(1-\frac{1}{3})}{3} + \frac{(1-\frac{1}{2})(1-\frac{1}{3})}{2^2.3} = \frac{5}{36}.$$

We now determine j_2. The set of numbers of the form $2^{n_1} 3^{n_2}$ has density zero. Thus the complementary set C_1 has density 1. We now take numbers 5, 7, 10, 11, 13, 14, 15, 17, ..., M_2 of C_1 so that

$$\frac{\sum \frac{1}{n} \text{ of these number}}{\sum\limits_{n=1}^{M_2} \frac{1}{n}} > \frac{9}{10}.$$

Let
$$\phi(m) = (1-1/q_1) \dots (1-1/q_m)$$

We take a j_2 so large that

$$\beta_2 = \frac{5}{36} + \phi(j_2) \cdot \left\{ 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \dots + \frac{1}{M_2} \right\} < \frac{3}{4}.$$

We now introduce basic vectors so that 5, 7, 10, ..., M_2 all become members of our right-complete set. In order to admit 5, we declare $(0, 0, 1, 0, 0, \dots)$ as a basic vector. In order to admit 7, we declare $(0, 0, 0, 1, 0, 0, \dots)$ as a basic vector. For 10, we declare $(1, 0, 1, 0, 0, \dots)$, and proceed like this until M_2 gains entry into our right-complete set. Of course, we make j_2 so large that $q_{j_2} > M_2$.

Let C_2 be the complement of the set of numbers of the form $2^{n_1} \dots q_{j_2}^{n_{j_2}}$. We choose an M_3 so large that

$$\frac{\left\{ \sum\limits_{\substack{(j_2+1) \leqslant n \leqslant M_3, \\ n \varepsilon C_2}} \frac{1}{n} \right\}}{\sum\limits_{1}^{M_3} \frac{1}{n}} > \frac{10}{11}.$$

We choose a j_3 so large that $\beta_3 = \beta_2 + \phi(j_3) \left(1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \dots + \frac{1}{M_3} \right) < \frac{3}{4}$.

We then admit basic vectors so that all n in $\underset{n}{E} \left\{ n \varepsilon C_2, \; q_{(j_2+1)} \leqslant n \leqslant M_3 \right\}$ gain entry into our right-complete set. Proceeding like this, we construct a right-complete (with respect to J) set whose upper logarithmic density is $= 1$ but whose magnification has measure $< \frac{3}{4}$.

Now, let J be any given sequence such that $\frac{\log j_{n+1}}{\log j_n}$ is unbounded. The counter example given above can be modified so as to prove the 'only if' part, as follows. Suppose $q_{k+1}, \dots, q_l$ is a block of consecutive primes. Let M be such that $q_{k+1} < M < q_l$. Consider the set of numbers all of whose prime factors are exclusively from among $q_1, q_2, \dots, q_k$, let C_k be the complement of this set. Consider the quantity

$$\frac{\left\{ \sum\limits_{\substack{q_{k+1} \\ \leqslant n \leqslant M \\ n \varepsilon C_k}} \frac{1}{n} \right\}}{\sum\limits_{1}^{M} \frac{1}{n}} \geqslant \frac{\log M - e^{\nu} \log q_k}{\log M}.$$

approximately (ν denotes Euler's constant),

$$= 1 - e^\nu . \frac{\log q_k}{\log M}.$$

We now use the following lemma :

Lemma :   *Let $a_1, a_2, \ldots$ be increasing sequence of positive integers such that $\frac{\log a_{n+1}}{\log a_n}$ is unbounded.   Take any $\epsilon > 0, \delta > 0$.   We can determine an n and a positive, integer M such that*

$$a_n < M < a_{n+1} \quad and \quad \frac{\log a_n}{\log M} < \epsilon \quad and \quad \frac{\log M}{\log a_{n+1}} < \delta.$$

Rigorizing the nonrigorous part above is trivial.

Corollary :   *Let $f(n)$ be a finite real-valued function defined on the set of positive integers.   Suppose there is a sequence J of positive integers j_n such that $\frac{\log j_{n+1}}{\log j_n}$ is bounded and $f(2^{x_1} \ldots q_{j_n}^{x_{j_n}})$ converges with probability 1 to a random variable $g(x)$, as $n \to \infty$.   Then f has a distribution and this is the same as the distribution of $g(x)$; we use logarithmic density.*

REFERENCE

PAUL, E. M. (1962) :   Density in the light of probability theory.   *Sankhyā*, Series A, 24, 103–114.

LIMIT THEOREMS FOR SUMS OF INDEPENDENT RANDOM VARIABLES WITH VALUES IN A HILBERT SPACE

By S. R. S. VARADHAN

Indian Statistical Institute

SUMMARY. In this article distributions on a Real Separable Hilbert Space are considered. Limit distributions are derived for sums of infinitesimal random variables.

A representation similar to the Levy-Khintchine representation is derived for infinitely divisible distributions. Necessary and sufficient conditions for compactness are obtained in terms of the quantities occurring in the representation.

1. INTRODUCTION

The celebrated Levy-Khintchine representation for the characteristic function $\phi(t)$ of an infinitely divisible distribution on the real line is given by

$$\phi(t) = \exp\left[i\nu t - \frac{\sigma^2 t^2}{2} + \int_{-\infty}^{\infty} \left(e^{itx} - 1 - \frac{itx}{1+x^2} \right) \frac{1+x^2}{x^2} \, dG(x) \right] \quad \dots \quad (1.1)$$

where ν and σ are real constants, $\sigma \geqslant 0$ and $G(x)$ is a bounded non-decreasing function of x which is continuous at the origin. ν, σ and G are uniquely determined by $\phi(t)$. Conversely, any function of the type (1.1) is the characteristic function of an infinitely divisible distribution. Khintchine and Bawly went further and proved that the limit distributions of sums of uniformly infinitesimal random variables are infinitely divisible and that they can be obtained as limits of certain accompanying infinitely divisible distributions. For all historical and other details concerning these we refer to Gnedenko and Kolmogorov (1954). Attempts have been made by several authors to extend these results to more general situations. We mention, in particular, the works of Bochner (1955) and (1958), Hunt (1956), Kloss (1961), Levy (1939), Takano (1955) and Vorobev (1954). Takano has generalised both the representation and the limit theorems to the case of a finite dimensional vector space. Levy has considered the circle group (the multiplicative group of complex numbers of modulus unity). The axiomatic development of the concept of characteristic functions by Bochner (1958) throws some light on the nature of the problem. Hunt has obtained the representation for one parametric semigroups of distributions on a Lie group. Vorobev has considered finite groups and Kloss a wide class of compact groups.

The case of a general locally compact abelian group was considered very recently by K. R. Parthasarathy, R. Rangarao and the present author (1962a). The results obtained are mentioned briefly in the next section and we will have occasion to use them in the later sections.

The case of a non-locally compact group offers considerable difficulty. However, the particular case of a separable Hilbert space can be tackled using some compactness criteria of Prohorov (1956). In this paper we will consider the case of a Hilbert space, and the analysis will throw some light on the precise nature of the difficulties in extending these results to non-locally compact groups.

2. The locally compact case

In the present section we will briefly mention the results obtained for the locally compact case. We will be using some of these results in the latter sections.

Let X be a locally compact abelian group which is separable metric and let Y be its character group. For $x \epsilon X$ and $y \epsilon Y$ we write (x, y) to denote the value of the character y at the point x.

Let $\mathcal{M}$ denote the convolution semi-group of probability measures on X, endowed with the weak topology. A sequence μ_n in $\mathcal{M}$ is shift compact whenever the sequence formed by some translates of these distributions is compact. Throughout the article ∗ denotes the convolution operation. When x is an element of the group and λ a distribution $\lambda * x$ denotes the convolution of λ with the distribution degenerate at x. λ^n denotes the n-th convolution of λ. $|\lambda|^2$ denotes $\lambda * \bar{\lambda}$ where $\bar{\lambda}$ is the distribution defined by $\bar{\lambda}(A) = \lambda(-A)$. $-A$ consists of the set of inverses of A. The following theorem was proved by Parthasarathy et al (1962).

Theorem 2.1 : *Let $\mu_n = \alpha_n * \beta_n$ for each n. If μ_n is compact then α_n and β_n are shift compact.*

Definition 2.1 : A distribution μ is said to be infinitely divisible if for each n there are elements x_n in X and $\lambda_n \epsilon \mathcal{M}$ such that

$$\mu = \lambda_n^n * x_n.$$

Remark 2.1 : This modification of the classical definition is intended to avoid the consequence of the presence of elements that are not divisible. In a divisible group this definition is equivalent to the classical definition as can be easily seen.

Theorem 2.2 : *The totality of infinitely divisible distributions is a closed sub-semigroup of $\mathcal{M}$.*

Definition 2.2 : If F is any totally finite measure on X the distribution $e(F)$ associated with it is defined as follows.

$$e(F) = e^{-F(X)} \left[1 + F + \frac{F^2}{2!} + \dots \right].$$

For $\mu \epsilon \mathcal{M}$ we denote by $\mu(y)$ its characteristic function. We then have

$$e(F)(y) = [\int [(x, y) - 1] dF].$$

Theorem 2.3 : *If μ is infinitely divisible and if its characteristic function vanishes at some point then μ has an idempotent factor.*

Theorem 2.4 : *Let $\mu_n = e(F_n)$. Then the necessary and sufficient conditions that*

(a) *μ_n is shift compact,*

(b) *if μ is any limit of shifts of μ_n then μ has no idempotent factors, are*

(i) *for each neighbourhood N of identity e the family F_n restricted to $X - N$ is weakly conditionally compact;*

(ii) *for each $y \epsilon Y$*

$$\sup_n \int [1 - \text{Real } (x, y)] dF_n < \infty.$$

Definition 2.3 : A double sequence $\alpha_{nj} : j = 1, 2, ..., k_n;\ \ n = 1, 2, ...$ of distributions is said to be uniformly infinitesimal if

$$\underset{n \to \infty}{\text{lt}} \ \underset{1 \leqslant j \leqslant k_n}{\sup} \ \underset{y \varepsilon K}{\sup} \ |\alpha_{nj}(y) - 1| = 0$$

for every compact subset $K \subseteq Y$.

Lemma 2.1 : *There exists a real valued function $g(x, y)$ on $X \times Y$ which is continuous in both the variables x and y and such that*

(i) $g(x, y_1 + y_2) = g(x, y_1) + g(x, y_2)$ *for each $x \varepsilon X$ and $y_1, y_2 \varepsilon Y$,*

(ii) *for any compact subset $K \subseteq Y$*

$$\underset{x \varepsilon X}{\sup} \ \underset{y \varepsilon K}{\sup} \ |g(x, y)| < +\infty,$$

(iii) *for each compact subset $K \subseteq Y$ there exists a neighbourhood N_K of the identity in X such that*

$$(x, y) = e^{ig(x,y)}$$

holds for $x \in N_K$ and $y \varepsilon K$,

(iv) $g(x, y) = 0$ *whenever $x = e$ for any y.*

Theorem 2.5 : *Let α_{nj} be a uniformly infinitesimal sequence and let*

$$\mu_n = \prod_{j=1}^{k_n} \alpha_{nj}.$$

Suppose μ_n is shift compact and that no limit of shifts of μ_n has an idempotent factor. Let $\beta_{nj} = e[\alpha_{nj} * g_{nj}]$ where g_{nj} is that element of the group X defined by the equality

$$(g_{nj},\ y) = \exp\{-i \int g(x, y) d\alpha_{nj}\}.$$

If

$$\lambda_n = \prod_{j=1}^{k_n} \beta_{nj} * g_n \text{ where } g_n = - \sum_j g_{nj},$$

then

$$\overline{\underset{n \to \infty}{\lim}} \ \underset{y \varepsilon K}{\sup} \ |\lambda_n(y) - \mu_n(y)| = 0$$

for every compact set K of Y.

Corollary 2.1 : *Limit distribution of sums infinitesimal independent random variables is infinitely divisible.*

Definition 2.4 : A distribution μ is said to be Gaussian if it has the following properties,

(i) μ is infinitely divisible,

(ii) $\mu = e(F) * \alpha$, where α is infinitely divisible implies that F vanishes outside the identity.

Therem 2.6 : *A distribution μ is Gaussian if and only if $\mu(y)$ is of the form*
$$\mu(y) = (x, y) \exp[-\phi(y)]$$
where x is a fixed element of X and $\phi(y)$ a continuous non-negative function of y satisfying the equation
$$\phi(y_1 + y_2) + \phi(y_1 - y_2) = 2[\phi(y_1) + \phi(y_2)]$$

for every pair y_1, y_2 in Y.

- 215 -

2

Theorem 2.7 :   *A distribution μ on X is infinitely divisible without an idempotent factor if and only if $\mu(y)$ is of the form*

$$\mu(y) = (x_0, y) \exp \left[\int [(x, y) - 1 - ig(x, y)]dM - \phi(y) \right]$$

where

(i)   *x_0 is a fixed element of X,*

(ii)   *$\phi(y)$ is a function as in Theorem 2.6,*

(iii)   *M is a σ-finite measure giving finite mass outside every neighbourhood of the identity,*

(iv)   *$\int [1 - \text{Real}\,(x, y)]dM < +\infty$ for each $y \in Y$.*

Remark 2.2 :   This representation is not in general unique. Examples of non-uniqueness and conditions for uniqueness are discussed by Parthasarathy *et al* (1962a) . It turns out that if the character group is connected then the representation is unique.

Corollary 2.2 :   *Every one parametric weakly continuous semigroup μ_t has a unique representation*

$$\mu_t(y) = (x_t, y) \ \exp \left[t \int [(x, y) - 1 - ig(x, y)]dM - t\phi(y) \right]$$

where x_t is a continuous semigroup in X and M and ϕ are as in Theorem 2.7.

Remark 2.3 :   Of the results mentioned in this section Theorems 2.1, 2.2 and the necessity part of Theorem 2.3 are valid in any complete separable metric group. We will also need the following corollary and theorem deducible from Theorem 2.2.

Corollary 2.3 :   *λ_n is shift compact if and only if $|\lambda_n|^2$ is compact.*

Theorem 2.8 :   *Let λ_n be a sequence such that λ_n is a factor of λ_{n+1} for all n. Then if λ_n is shift compact there are translates λ_n' of λ_n which converge.*

Remark 2.4 :   We can assume instead that λ_{n+1} is a factor of λ_n for each n and then also the theorem will be valid.

3.   PRELIMINARIES

X is a real separable Hilbert space, (x, y) denotes the inner product and $\|x\|$ the norm. With vector addition as group operation X becomes a complete separable metric group. We will denote by $\mathcal{M}$ the semigroup of all distributions.

For every $\mu \in \mathcal{M}$ its characteristic function is defined on X by the formula

$$\mu(y) = \int e^{i(x,\, y)} \ d\mu(x).$$

We will, in this section, mention some results obtained by Prohorov (1956) concerning compactness criteria for distributions on X.

Definition 3.1 :   A positive semi-definite Hermitian operator A is called an S-operator if it has finite trace. The class of sets of the type $[x : (Sx, x) < t]$ where S runs over S-operators and t over positive numbers forms a neighbourhood system at the origin for a certain topology which is called the S-topology. A net x_α converges to zero in S-topology if and only if (Sx_α, x_α) converges to zero for every S-operator S.

We have the following theorem concerning characteristic functions and S-topology which was obtained by Sazanov (1958).

Theorem 3.1 : *In order that a function $\mu(y)$ may be the characteristic function of a distribution on X it is necessary and sufficient that $\mu(0) = 1$, $\mu(y)$ be positive definite and continuous at $y = 0$ in the S-topology. (Here and elsewhere in the article 0 will denote the null element of X or the identity of the group and will be called the origin).*

We also have the following theorem of Prohorov (1956).

Theorem 3.2 : *In order that a positive definite function $\mu(y)$ with $\mu(0) = 1$ be the characteristic function of a distribution on X it is necessary and sufficient that for every $\epsilon > 0$ there exists an S-operator S_ϵ such that*

$$1 - R\mu(y) \leqslant (S_\epsilon y, y) + \epsilon$$

where R denotes the real parts.

Definition 3.2 : Let μ be a distribution on X for which $\int (x, y)^2 d\mu < \infty$ for each y. Then the covariance operator S of μ is that Hermitian operator for which

$$(Sy, y) = \int (x, y)^2 d\mu(x).$$

This operator S will be positive semi-definite and will be an S-operator if and only if

$$\int \|x\|^2 \, d\mu(x) < \infty.$$

Definition 3.3 : A sequence S_n of S-operators will be called compact if and only if the following two conditions are satisfied

$$\text{(i)} \qquad \sup_n \text{Trace-}(S_n) < \infty$$

$$\text{(ii)} \qquad \lim_{N \to \infty} \sup_n \sum_{j=N}^{\infty} (S_n e_j, e_j) \to 0$$

for some orthonormal sequence $e_1, e_2, \dots e_j, \dots$

When S is the covariance operator of a distribution μ on X for which

$$\int \|x\|^2 d\mu(x) < \infty$$

we have the relation
$$\sum_{j=N}^{\infty} (Se_j, e_j) = \int r_N^2(x) d\mu(x)$$

where
$$r_N^2(x) = \sum_{j=N}^{\infty} (x, e_j)^2$$

and $e_1, e_2, \dots$ any orthonormal basis.

The following theorems concerning conditions for compactness of a sequence μ_n of distributions on X were obtained by Prohorov (1956).

Theorem 3.3 : *In order that a sequence μ_n of distributions on X be weakly conditionally compact it is necessary and sufficient that for every $\epsilon > 0$, there exists a compact sequence S_n^ϵ of S-operators such that*

$$1 - R \mu_n(y) \leqslant (S_n^\epsilon y, y) + \epsilon$$

for all n and y. Here $\mu_n(y)$ is the characteristic function of μ_n.

Theorem 3.4 : *Let μ_n be a sequence of distributions for which the covariance operators S_n exist and are S-operators. Let further S_n be compact. Then μ_n is weakly conditionally compact.*

Theorem 3.5 : *In order that a sequence μ_n of distributions on X may be weakly conditionally compact it is necessary that for any $\epsilon > 0$*

$$\lim_{N \to \infty} \sup_n \mu_n[r_N^2(X) > \epsilon] = 0.$$

In defining the covariance operator in Definition 3.2. we assumed that μ is a distribution. Actually, if M is any σ-finite measure for which

$$\int \|x\|^2 \, dM < \infty$$

then S is well defined as an S-operator by the relation

$$(Sy, y) = \int (x, y)^2 \, dM(x).$$

We will need also the following.

Theorem 3.6 : *(Rangarao, 1960). Let $\mu_n \to \mu$. Let $\{f_\alpha\}$ be a family of continuous functions equicontinuous and uniformly bounded. Then*

$$\lim_{n \to \infty} \sup_\alpha \left| \int f_\alpha d\mu_n - \int f_\alpha d\mu \right| = 0.$$

4. An estimate of the variance

Let $X_1, X_2, \ldots, X_n$, be n symmetric independent random variables in the Hilbert space X. We will give in this section an estimate for the variance $E\|X_1 + \ldots X_n\|^2$ when each X_j is bounded uniformly in norm by a constant C independent of j.

To this end we introduce the concentration function $Q_\mu(t)$ following Levy [(1954), page 138].

Definition 4.1 : The concentration function $Q_\mu(t)$ of a distribution μ in the Hilbert space X is defined for $0 < t < \infty$ as

$$Q_\mu(t) = \sup_{x \in X} \mu(S_t + x)$$

where S_t denotes the sphere $[x : \|x\| \leqslant t]$ and $S_t + x$ its translate by the element x of X.

We now list a few elementary properties of these functions.

Theorem 4.1 : (1) *$Q_\mu(t)$ is a nondecreasing function of t and $\lim_{t \to \infty} Q_\mu(t) = 1$.* (2) *If $\mu_1 * \mu_2 = \mu$ then, for every t, $Q_\mu(t) \leqslant \min [Q_{\mu_1}(t), Q_{\mu_2}(t)]$.* (3) *If μ_n is shift compact then*

$$\lim_{t \to \infty} \inf_n Q_{\mu_n}(t) = 1.$$

Proof : (1) and (2) are obvious and (3) is a consequence of tightness.

Theorem 4.2 : *Let $X_1, X_2, \ldots, X_n$ be n mutually independent symmetric random variables. Let $S_j = X_1 + \ldots + X_j$. Further, let $Q(t)$ denote the concentration function of the sum $S_n = X_1 + \ldots X_n$. If T is defined as*

$$T = \sup_{1 \leqslant j \leqslant n} \|S_j\|$$

then one has for any $t > 0$ $\quad P\{T > 4t\} \leqslant 2[1 - Q(t)].$

Proof : This remarkable result of Levy, which is a refinement of Kolmogorov's inequality in terms of the variance, although proved by him for the real line, offers no difficulty for generalisation. Let the events E_k be defined as follows :

$$E_k = [\|S_1\| \leqslant 4t, ..., \|S_{k-1}\| \leqslant 4t, \|S_k\| > 4t]$$
$$\{T > 4t\} = E_1 \cup E_2 ... \cup E_n, \quad E_i \cap E_j = \phi.$$

By $Pr\{\ \}$ we will denote the probability of the event within the brackets given that the event E_r has occurred. We then have

$$Pr\{\|S_n\| \leqslant 2t\} \leqslant Pr\{\|S_n - S_r\| > 2t\} = P\{\|S_n - S_r\| > 2t\}. \qquad ... (4.1)$$

This is because E_r and $\|S_n\| \leqslant 2t$ imply that $\|S_n - S_r\| > 2t$ and $S_n - S_r$ is distributed independently of E_r. Let us further suppose that $Q(t) > \frac{1}{2}$. We now consider the distribution μ_{rn} of $S_n - S_r$ whose concentration function is denoted by $Q_{rn}(t)$. Since μ_{rn} is a factor of μ we have $Q_{rn}(t) > \frac{1}{2}$. This implies the existence of a point x in the space X such that

$$\mu_{rn}(S_t + x) > Q_{rn}(t) - \epsilon > \frac{1}{2}. \qquad ... (4.2)$$

Since μ_{rn} is a symmetric distribution

$$\mu_{rn}(S_t - x) = \mu_{rn}(-S_t + x) = \mu_{rn}(S_t + x) > \frac{1}{2}.$$

Therefore,

$$S_t - x \cap S_t + x \neq \phi.$$

In other words there exists a point y such that

$$\|x + y\| \leqslant t, \quad \|x - y\| \leqslant t.$$

These two imply that $\|x\| \leqslant t$ and hence

$$x + S_t \subseteq S_{2t}. \qquad ... (4.3)$$

From (4.2) and (4.3) follows $\quad \mu_{rn}(S_{2t}) > Q_{rn}(t) - \epsilon.$

Since ϵ is arbitrary we have $\quad \mu_{rn}(S_{2t}) \geqslant Q_{rn}(t) \qquad ... (4.4)$

which is the same as $\quad P\{\|S_n - S_r\| > 2t\} \leqslant 1 - Q_{rn}(t) \leqslant 1 - Q(t). \qquad ... (4.5)$

We have further, using (4.1) and (4.5)

$$P\{T > 4t, \|S_n\| \leqslant 2t\} = \sum_{r=1}^{n} Pr\{\|S_n\| \leqslant 2t\} P[E_r] \leqslant (\sum_{r=1}^{n} P[E_r](1 - Q(t)) = P\{T > 4t\}(1 - Q(t)),$$
$$... (4.6)$$

we also have $\quad P\{T \leqslant 4t, \|S_n\| \leqslant 2t\} \leqslant P\{T \leqslant 4t\} = 1 - P\{T > 4t\}. \qquad ... (4.7)$

Adding (4.6) and (4.7) we get $\quad P\{\|S_n\| \leqslant 2t\} \leqslant 1 - P\{T > 4t\} Q(t). \qquad ... (4.8)$

From (4.4) putting $r = 0$, we obtain

$$P\{\|S_n\| \leqslant 2t\} \geqslant Q(t), \qquad ... (4.9)$$

(4.8) and (4.9) imply, since $Q(t) > \frac{1}{2}$,

$$P\{T > 4t\} \leqslant \frac{1 - Q(t)}{Q(t)} \leqslant 2[1 - Q(t)]. \qquad ... (4.10)$$

However, if $Q(t) \leqslant \frac{1}{2}$ the theorem is trivially true. Thus the proof of the theorem is complete.

We have the following theorem which gives an estimate of the variance. This is well known for the real line. We reproduce the proof for the real line, which can be found in Halmos [(1950) page 198], replacing however $|x|$ by $\|x\|$.

Theorem 4.3 : *Let $X_1, X_2, ..., X_n$ be n independent random variables in a Hilbert space, uniformly bounded by a constant c in norm. Let each X_i have zero expectation. In addition let*

$$P\{ \sup_{1 < j < n} \|S_j\| \leqslant d\} \geqslant \epsilon > 0$$

where

$$S_j = X_1 + ... X_j$$

Then

$$E \|S_n\|^2 \leqslant \frac{d^2 + (c+d)^2}{\epsilon} .$$

Proof : Let the sets E_k be defined as follows

$$E_k = \sup_{1 < j < k} \|S_j\| \leqslant d.$$

Then $E_1 \supset E_2 ... \supset E_n$ and $P(E_n) \geqslant \epsilon > 0$. By P we mean the product measure of $X_1, X_2, ..., X_n$. Let us define

$$F_k = E_{k-1} - E_k$$

and

$$\alpha_k = \int_{E_k} \|S_k\|^2 \, dP.$$

We will take E_0 as the whole space and α_0 as zero. We then have

$$\alpha_k - \alpha_{k-1} = \int_{E_k} \|S_k\|^2 \, dP - \int_{E_{k-1}} \|S_{k-1}\|^2 \, dP$$

$$= \int_{E_{k-1}} \|S_k\|^2 \, dP - \int_{F_k} \|S_k\|^2 \, dP - \int_{E_{k-1}} \|S_{k-1}\|^2 \, dP$$

$$= \int_{E_{k-1}} \|S_{k-1}\|^2 \, dP - 2 \int_{E_{k-1}} (S_{k-1}, X_k) \, dP + \int_{E_{k-1}} \|X_k\|^2 \, dP$$

$$- \int_{F_k} \|S_k\|^2 \, dP - \int_{E_{k-1}} \|S_{k-1}\|^2 \, dP$$

$$= \int_{E_{k-1}} \|X_k\|^2 \, dP - \int_{F_k} \|S_k\|^2 \, dP$$

$$\geqslant P(E_{k-1}) E \|X_k\|^2 - (c+d)^2 P(F_k). \qquad \qquad .... (4.11)$$

We notice here that $\int_{E_{k-1}} (S_{k-1}, X_k) \, dP = 0$ since X_k is independent of S_{k-1}, has zero expectation and is independent of E_{k-1} as well. Moreover, $F_k \subseteq E_{k-1}$ hence $\|S_k\| \leqslant \|S_{k-1}\| + \|X_k\| \leqslant (c+d)$ over F_k. Since $P(E_n) \leqslant P(E_k)$ for any k we have

$$\alpha_k - \alpha_{k-1} \geqslant P(E) E \|X_k\|^2 - (c+d)^2 P(F_k) \quad \text{for } k = 1, 2, ... \qquad ... (4.12)$$

Since $F_1, F_2, ..., F_n$ are all disjoint adding (4.12) for $k = 1, 2, ..., n$, we have

$$\alpha_n \geqslant P(E) E \|S_n\|^2 - (c+d)^2. \qquad \qquad ... (4.13)$$

However, since $\|S_n\| \leqslant d$ on E_n $\alpha_n = \int_{E_n} \|S_n\|^2 \, dP \leqslant d^2,$ $\qquad \qquad ... (4.14)$

(4.13) and (4.14) give

$$E\|S_n\|^2 \leqslant \frac{(d+d)^2 + d^2}{P(E)} \leqslant \frac{(c+d)^2 + d^2}{\epsilon} .$$

This completes the proof.

Theorem 4.4 : *Let X_1, X_2, ..., X_n be symmetric independent random variables in the Hilbert space such that $\| X_i \| \leqslant c$ for $i = 1, 2, ..., n$. Let $Q(t)$ denote the concentration function of $S_n = X_1 + ... X_n$. Then*

$$E\|S_n\|^2 \leqslant \frac{16t^2 + (c + 4t)^2}{2Q(t) - 1}$$

for any t such that $Q(t) > \frac{1}{2}$.

Proof : This follows at once from Theorems 4.2 and 4.3 and the fact that a bounded symmetric random variable has zero expectation.

5. INFINITELY DIVISIBLE DISTRIBUTIONS

In this section we will obtain a representation for infinitely divisible distributions. The definition of an infinitely divisible distribution is the same as Definition 2.1, but since the Hilbert space as a group is divisible it is equivalent to the classical definition which requires that μ be written as λ_n^n for each n.

As we have already remarked at the end of Section 2 that some of the results mentioned in that Section are valid for the Hilbert space we will now state them. We will keep in mind also that the Hilbert space has no nontrivial compact subgroups and hence there are no idempotent distributions.

Theorem 5.1 : *The infinitely divisible distributions form a closed sub-semigroup among all distributions.*

Theorem 5.2 : *If μ is an infinitely divisible distribution and $\mu(y)$ its characteristic function then $\mu(y)$ is nonvanishing.*

For every finite measure F the infinitely divisible distribution $e(F)$ is associated in the same way as in Definition 2.2. We then have the following theorem.

Theorem 5.3 : *Let $\mu_n = e(F_n)$. In order that μ_n may be shift compact it is necessary that*

(i) *for each neighbourhood N of the identity F_n restricted to N' is weakly conditionally compact,*

(ii) $\sup_n \int [1 - \cos(x, y)] dF_n < \infty$ *for each y.*

Theorem 5.4 : *Let $\mu_n \Longrightarrow \mu$. Then $\mu_n(y) \to \mu(y)$ uniformly over every bounded sphere.*

Proof : Since the set of functions (x, y) as y varies over a bounded sphere forms a equicontinuous family of functions in x the theorem follows.

Theorem 5.5 : *Let μ_n be shift compact and $\mu_n(y) \to \mu(y)$ uniformly over bounded spheres. Then $\mu_n \Longrightarrow \mu$.*

Proof : Since μ_n is shift compact let x_n in X be chosen such that $\mu_n * x_n$ is compact. We will now show that x_n in X is compact. If x_n is not compact then we can produce a subsequence from x_n which has no further convergent subsequence. We will denote the subsequence by x_n itself. Since $\mu_n * x_n$ is weakly compact it has

a subsequence $\mu_{n_j} * x_{n_j}$ converging weakly. Thus $\mu_{n_j}(y)e^{i(x_{n_j}, y)}$ as well as $\mu_{n_j}(y)$ converge uniformly over bounded spheres. Since from the compactness of $|\mu_n|^2$ one can conclude the existence of a sphere S such that $\inf |\mu_n(y)| \geqslant \epsilon > 0$ for all y in S, it follows that $e^{i(x_{n_j}, y)}$ converges uniformly over S and hence x_{n_j} converges in norm. This proves the theorem.

Before obtaining the representation we will show that if $e(F_n)$ is shift compact then

$$\sup_n \int_{\|x\| < 1} \|x^2\| dF_n < \infty.$$

To this end we consider the following lemma.

Lemma 5.1 : *Let $f(y)$ be a non-negative function on X such that $f(2y) \leqslant 4f(y)$ for all values of y. If $f(y) \leqslant \epsilon$ whenever $(Sy, y) \leqslant \delta$ where S is some S-operator then*

$$f(y) \leqslant (S_1 y, y) + \epsilon \quad \text{for all } y$$

where

$$S_1 = 4\epsilon\delta^{-1}S.$$

Proof : Defining $S_0 = \epsilon\delta^{-1}S$,

we see that

$$f(y) \leqslant \epsilon \quad \text{whenever} \quad (S_0 y, y) \leqslant \epsilon. \qquad \ldots \text{(5.1)}$$

Further, if $(S_0 y, y) \leqslant 4^n\epsilon$ where n is a positive integer, then denoting by y_n the element $2^{-n} y$, we have

$$(S_0 y_n, y_n) = 4^{-n}(S_0 y, y) \leqslant \epsilon$$

consequently $f(y_n) \leqslant \epsilon$. But since $f(2y) \leqslant 4f(y)$,

we have

$$f(y) \leqslant 4^n f(y_n) < 4^n\epsilon.$$

So from (5.1) we have

$$f(y) \leqslant 4^n\epsilon \quad \text{whenever} \quad (S_0 y, y) \leqslant 4^n\epsilon. \qquad \ldots \text{(5.2)}$$

Let y be any element of X and let $(S_0 y, y) = t$.

Case (i) : Let $t > \epsilon$. If n is a non-negative integer such that

$$4^n \epsilon < t \leqslant 4^{n+1}\epsilon. \qquad \ldots \text{(5.3)}$$

We have, since $(S_0 y, y) = t \leqslant 4^{n+1}\epsilon$, using (5.2) and (5.3)

$$f(y) \leqslant 4^{n+1} \epsilon \leqslant 4t = 4(S_0 y, y). \qquad \ldots \text{(5.4)}$$

Case (ii) : Let $t \leqslant \epsilon$. Then from (4.1) we have

$$f(y) \leqslant \epsilon. \qquad \ldots \text{(5.5)}$$

(5.4) and (5.5) give at once $\quad f(y) \leqslant \max[\epsilon, (4S_0 y, y)] \leqslant \epsilon + (S_1 y, y)$.

We shall need while proving the next theorem the following inequality. If $a_1, \ldots a_m$ are any m real numbers such that $|a_j| \leqslant 1$ for $1 \leqslant j \leqslant m$, then

$$1 - a_1 a_2 \ldots a_m \leqslant \sum_{j=1}^{m} (1 - a_j). \qquad \ldots \text{(5.6)}$$

This inequality is proved by induction if all the a's are positive. If at least one of them say a_r is negative

$$1 - a_1 \ldots a_m \leqslant 1 + |a_1 \ldots a_m| \leqslant 1 + |a_r| = 1 - a_r \leqslant \sum_{j=1}^{m} (1 - a_j).$$

We will now prove the following theorem.

Theorem 5.6 : *Let F_n be a sequence of finite measures such that $e(F_n)$ is shift compact. Then*

$$\sup_{n} \int_{\|x\| \leqslant 1} \|x\|^2 \, dF_n < \infty.$$

Proof : We assume without any loss of generality that each F_n vanishes outside the unit sphere. Otherwise we can consider the restriction of F_n to the unit sphere instead of F_n. Let $M_n = F_n + \bar{F}_n$. Then $e(M_n) = |e(F_n)|^2 = \lambda_n$ is compact. We will show that $\int \|x\|^2 \, dM_n$ is uniformly bounded. To this end we assume that the total mass of M_n is an integer for every n. If this were not so we can write $M_n = M_n^{(1)} + M_n^{(2)}$ where $M_n^{(1)}$ is symmetric with an integral total mass and $M_n^{(2)}$ has total mass less than unity. Consequently

$$\int_{\|x\| \leqslant 1} \|x\|^2 \, dM_n^{(2)} \leqslant 1.$$

Since our aim is to prove that $\quad \sup_{n} \int_{\|x\| \leqslant 1} \|x\|^2 \, dM_n < \infty$

it suffices to show that $\quad \sup_{n} \int_{\|x\| \leqslant 1} \|x\|^2 \, dM_n^{(1)} < \infty$.

Now since the total mass of M_n is an integer say k_n we will write

$$M_n = F_{n1} + \ldots F_{nk_n}$$

where F_{nj} for $j = 1, 2, \ldots, k_n$, $n = 1, 2, \ldots$ is a symmetric probability measure in the unit sphere. Let us now denote by μ_n the convolution

$$\mu_n = F_{n1} * F_{n2} \ldots * F_{nk_n}.$$

Since each F_{nj} is symmetric and has zero expectation

$$\int \|x\|^2 \, d\mu_n = \sum_{j=1}^{k_n} \int \|x\|^2 \, dF_{nj} = \int \|x\|^2 \, dM_n.$$

Hence it suffices to show that $\quad \sup_{n} \int \|x\|^2 \, d\mu_n < \infty$.

If $Q_n(t)$ denotes the concentration function of μ_n from Theorem 4.4 we have

$$\int \|x\|^2 \, d\mu_n \leqslant \frac{16t^2 + (4t+1)^2}{2Q_n(t) - 1}$$

whenever $Q_n(t) > \frac{1}{2}$. Therefore, it is enough to prove that

$$\inf_{n} Q_n(t) \geqslant \tfrac{3}{4} \quad \text{for some } t$$

which will follow from Theorem 4.1 if we prove that μ_n is weakly conditionally compact.

Since each $F_{nj}(y)$ is a real characteristic function and

$$\mu_n(y) = \prod_{j=1}^{k_n} F_{nj}(y) \qquad \qquad \ldots \text{(5.7)}$$

it follows from (5.6) that $\quad 1 - \mu_n(y) \leqslant \sum_{j=1}^{k_n} [1 - F_{nj}(y)] = \sum_{j=1}^{k_n} \int [1 - \cos(x, y)] \, dF_{nj}(x)$

$$= \int [1 - \cos(x, y)] \, dM_n(x) = f_n(y) \text{ say.} \qquad \ldots \text{(5.8)}$$

223

3

We also have $\lambda_n(y) = \exp[-f_n(y)]$ and hence for any given ϵ there exists a δ depending only on ϵ such that

$$f_n(y) \leqslant \epsilon \quad \text{if} \quad 1 - \lambda_n(y) \leqslant \delta. \qquad \dots (5.9)$$

Since λ_n is compact we have from Theorem 3.5 for any given $\delta > 0$ a compact sequence S_n of S-operators depending on δ only such that

$$1 - \lambda_n(y) \leqslant (S_n y, y) + \frac{\delta}{2}. \qquad \dots (5.10)$$

From (5.9) and (5.10) it follows that whenever $(S_n y, y) \leqslant \delta/2, f_n(y) \leqslant \epsilon$ and hence from Lemma 5.1 we have

$$f_n(y) \leqslant (S'_n y, y) + \epsilon \qquad \dots (5.11)$$

where
$$S'_n = 8\epsilon\delta^{-1} S_n.$$

Since S_n is compact so is S'_n and Theorem 3.3 and (5.11) imply that μ_n is weakly conditionally compact. The proof of the theorem is now complete.

We will denote by $K(x, y)$ the following function

$$K(x, y) = e^{i(x, y)} - 1 - \frac{i(x, y)}{1 + \|x\|^2}.$$

Theorem 5.7 : *Let μ_n for each n be of the form $e(F_n)$ where F_n is a finite measure. Let $\mu_n * x_n \Longrightarrow \mu$ for some suitably chosen elements x_n in X. We further assume that F_n is increasing. Then F_n increases to a measure F which may be σ-finite but gives finite mass outside every neighbourhood of the origin and for which*

$$\int\limits_{\|x\| \leqslant 1} \|x\|^2 dF < \infty.$$

In addition
$$\mu(y) = \exp[i(x_0, y) + \int K(x, y) dM(x)]$$
where x_0 is a fixed element of X.

Proof : Let $\lambda_n(y)$ be defined as

$$\lambda_n(y) = \exp[\int K(x, y) dF_n].$$

Then $\lambda_n(y)$ is the characteristic function of λ_n which is the shift of μ_n by the element

$$Z_n = -\int \frac{x}{1 + \|x\|^2} dF_n.$$

We will now show that $\lambda_n(y)$ converges uniformly in y over bounded spheres. For this purpose we write

$$\int K(x, y) dF_n = \int\limits_{\|x\| \leqslant 1} K(x, y) dF_n + \int\limits_{\|x\| > 1} K(x, y) dF_n. \qquad \dots (5.12)$$

Let F be the limit of F_n. From Theorems 5.3 and 5.6 it follows that F is finite outside every neighbourhood of the origin and

$$\int\limits_{\|x\| \leqslant 1} \|x\|^2 dF < \infty.$$

Since
$$|K(x, y)| \leqslant 2 + \frac{\|x\| \, \|y\|}{1 + \|x\|^2} \leqslant 2 + \|y\|$$

it follows that

$$\sup_{y \epsilon S} \left| \int_{||x|| > 1} K(x, y) dF_n - \int_{||x|| > 1} K(x, y) dF \right| \to 0 \text{ as } n \to \infty \quad \dots \text{ (5.13)}$$

for every bounded sphere S. On the other hand if $\|x\| \leqslant 1$

$$|K(x, y)| \leqslant |e^{i(x,y)} - 1 - i(x, y)| + \frac{|(x, y)| \, \|x\|^2}{1 + \|x\|^2}$$

$$\leqslant \tfrac{1}{2}(x, y)^2 + |(x, y)| \, \|x\|^2 \leqslant \tfrac{1}{2} \|x\|^2 \|y\|^2 + \|y\| \, \|x\|^2.$$

Since $\int_{||x|| \leqslant 1} \|x\|^2 \, dF$ is finite it follows that

$$\sup_{y \epsilon S} \left| \int_{||x|| \leqslant 1} K(x, y) dF_n - \int_{||x|| \leqslant 1} K(x, y) dF \right| \to 0 \text{ as } n \to \infty \quad \dots \text{ (5.14)}$$

for every bounded sphere S. (5.12), (5.13) and (5.14) imply that

$$\sup_{y \epsilon S} |\lambda_n(y) - \lambda(y)| \to 0 \quad \text{as } n \to \infty$$

for every bounded sphere S where $\lambda(y) = \exp [\int K(x, y) dF]$.

Since λ_n is shift compact from Theorem 5.5 it follows that $\lambda_n \Longrightarrow \lambda$ and λ has to be a shift of μ. Hence the theorem follows.

Theorem 5.8 : *Let $\mu(y)$ be a function of the form*

$$\mu(y) = \exp [\int K(x, y) dF(x)]$$

where F is a σ-finite measure giving finite mass outside every neighbourhood of the identity and for which

$$\int_{||x|| \leqslant 1} \|x\|^2 \, dF < \infty.$$

Then $\mu(y)$ is the characteristic function of an infinitely divisible distribution.

Proof : Let N_n denote the sphere of radius $\dfrac{1}{n}$ around the origin and F_n the restriction of F to N_n'. Then F_n increases to F. Let

$$\mu_n(y) = \exp [\int K(x, y) dF_n].$$

From the proof of Theorem 5.1 it follows that

$$\sup_{y \, \epsilon \, V} |\mu_n(y) - \mu(y)| \to 0 \text{ as } n \to \infty$$

for every bounded sphere V. In view of Theorems 5.1 and 5.5 it is enough to show that μ_n is shift compact. We will now show that $\lambda_n = |\mu_n|^2$ is compact.

$$\lambda_n = |\mu_n|^2 = |e(F_n)|^2 = e(M_n) \text{ where } M_n = F_n + \bar{F}_n.$$

Since F_n increases to F it follows that M_n increases to M where $M = F + \bar{F}$. Without any loss of generality we can assume that F and hence M vanishes outside the sphere $\|x\| \leqslant 1$. We further have

$$1 - \lambda_n(y) = 1 - \exp [\int [\cos (x, y) - 1] dM_n] = 1 - \exp [- \int [1 - \cos (x, y)] dM_n]$$

$$\leqslant \int [1 - \cos (x, y)] dM_n \leqslant \int [1 - \cos (x, y)] \, dM$$

$$\leqslant \tfrac{1}{2} \int_{||x|| < 1} (x, y)^2 \, dM = \tfrac{1}{2}(Sy, y).$$

Since

$$\int_{\|x\| \leqslant 1} \|x\|^2 \, dM = 2 \int_{\|x\| \leqslant 1} \|x\|^2 \, dF < \infty$$

it follows that S is an S-operator. Since S is a fixed S-operator independent of n it follows from Theorem 3.5 that λ_n is compact. Consequently μ_n is shift compact and the theorem is proved.

Gaussian distributions are defined in the Hilbert Space in exactly the same manner as in Definition 2.4. We shall now prove

Theorem 5.9 : *A distribution μ on X is Gaussian if and only if $\mu(y)$ is of the form*

$$\mu(y) = \exp \{i(x_0, y) - (Sy, y)\}$$

where X_0 is a fixed element and S an S-operator.

Proof : Let us take a countable dense subset $y_1, y_2, \ldots, y_n \ldots$ in X and consider the map τ from X to Z^∞, the countable product of the circle groups, defined as

$$\tau(x) = [e^{i(x, y_1)}, \ldots e^{i(x, y_n)}, \ldots].$$

Let H be the image of X under τ in Z^∞. Then τ is a both ways measurable isomorphism of the two groups X and H. If μ is Gaussian on X then $\mu\tau^{-1}$ is Gaussian in H and hence in Z^∞. From Theorem 2.5 we have

$$\mu\tau^{-1}(\theta) = \theta(Z) \exp[-\phi(\theta)] \qquad \qquad \ldots \ (5.15)$$

where θ is a character on Z^∞, Z is a fixed element of Z^∞ and ϕ a function with properties specified in Theorem 2.5. $\theta(Z)$ denotes the value of the character θ at the point Z. Further any character θ on Z^∞ is of the form

$$\theta(Z) = Z_1^{n_1} \ldots Z_k^{n_k} \qquad \qquad \ldots \ (5.16)$$

where n_1, n_k are integer and $Z_1, \ldots Z_k$ the first k coordinates of Z in Z^∞.

Therefore, $\mu\tau^{-1}(\theta) = \mu(n_1 y_1 + \ldots n_k y_k)$

where $\mu(y)$ is the characteristic function on X and $n_1, \ldots n_k$ are related to θ by the relation (5.16). Hence

$$|\mu\tau^{-1}(\theta)| = |\mu| (n_1 y_1 + \ldots n_k y_k) = e^{-\phi(\theta)}.$$

Since for any θ and θ'

$$\phi(\theta + \theta') + \phi(\theta - \theta') = 2[\phi(\theta) + \phi(\theta')]$$

it follows that

$$h(y + y') + h(y - y') = 2[h(y') + h(y)] \qquad \ldots \ (5.17)$$

whenever y, y' are of the form $n_1 y_1 + \ldots n_k y_k$, where

$$h(y) = -\log |\mu(y)|.$$

Since $y_1, y_2, \ldots$ are dense and h is continuous in the norm topology of X it follows that the relation (5.17) is valid for any pair y, y'. This implies that $h(y)$ can be put as

$$h(y) = (S\bar{y}, y)$$

where S is a positive semi-definite Hermitian operator. From the continuity of $\mu(y)$ in the S-topology it follows that S is an S-operator. If we now consider the distribution λ on X defined by the equation

$$\lambda(y) = \exp[-(Sy, y)].$$

We have

$$\lambda\tau^{-1}(\theta) = e^{-\phi(\theta)}.$$

Hence $\lambda\tau^{-1}$ is a shift of $\mu\tau^{-1}$. Since both $\lambda\tau^{-1}$ and $\mu\tau^{-1}$ give unit mass for the subgroup H of Z^∞ the element Z by which $\lambda\tau^{-1}$ is shifted to obtain $\mu\tau^{-1}$ belongs to H and hence $Z = \tau(x_0)$ for some $x_0 \epsilon X$. Consequently,

$$\mu = \lambda * x_0$$

or

$$\mu(y) = \exp[i(x_0, y) - (Sy, y)].$$

Conversely, if $\mu(y)$ is of the form $\quad \mu(y) = \exp[i(x_0, y) - (Sy, y)]$

the distribution $\mu\tau^{-1}$ in Z^∞ is Gaussian and hence so is μ.

We now prove the representation theorem for infinitely divisible distributions.

Theorem 5.10 : A function $\mu(y)$ is the characteristic function of an infinitely divisible distribution μ on X if and only if it is of the form

$$\mu(y) = \exp[i(x_0, y) - (Sy, y) + \int K(x, y)\, dM(x)] \qquad \cdots \quad (5.18)$$

where x_0 is a fixed element of X, S an S-operator and M a σ-finite measure giving finite-mass outside every neighbourhood of the origin and for which

$$\int_{\|x\| \leqslant 1} \|x\|^2\, dM < \infty.$$

Here $K(x, y)$ is the function $\quad K(x, y) = \left[e^{i(x,y)} - 1 - \dfrac{i(x, y)}{1 + \|x\|^2} \right].$

The representation (5.18) of $\mu(y)$ is unique.

Proof : Let $\mu(y)$ be the characteristic function of an infinitely divisible distribution μ. Then in the same manner as in the proof of Theorem 7.1 of Parthasarathy *et al* (1962a) we can construct a sequence of distributions λ_n such that (i) $\lambda_n = e(M_n)$, (ii) M_n increases to M, (iii) $\lambda_n * x_n \Longrightarrow \lambda$, (iv) $\mu = \lambda * \mu_0$ where μ_0 is Gaussian.

Now using Theorems 5.7 and 5.9 we can complete the proof. Sufficiency is immediate from Theorems 5.8 and 5.9. Uniqueness is proved in an exactly same manner as Theorem 8.1 of Parthasarathy *et al* (1962a) but keeping in mind that the space X playing the role of the character group is connected.

6. Compactness Criteria

In the present section we will find out the necessary and sufficient conditions in order that a sequence μ_n of infinitely divisible distributions may be weakly conditionally compact.

If μ is any infinitely divisible distribution by $\mu = [x, S, M]$ we will mean that the three quantities occurring in the representation of μ according to Theorem 5.10 are respectively x, S and M. We will associate with any such $\mu = [x, S, M]$ another S-operator which we will denote by T.

$$(Ty, y) = 2(Sy, y) + \int_{\|x\| \leqslant 1} (x, y)^2\, dM(x) \qquad \cdots \quad \cdots \quad (6.1)$$

T is an S-operator since $\quad \displaystyle\int_{\|x\| \leqslant 1} \|x\|^2\, dM(x) < \infty.$

Lemma 6.1 : *In order that a sequence μ_n of Gaussian distributions with covariance operators S_n be shift compact it is necessary and sufficient that S_n should be compact. [If μ is Gaussian with covariance operator S its characteristic function is* $\exp[i(x_0, y) - \frac{1}{2}(Sy, y)]$.

Proof : Sufficiency is immediate from Theorem 3.4. We will prove necessity. If μ_n is shift compact then $|\mu_n|^2$ is compact. But $|\mu_n|^2$ is Gaussian with mean zero and covariance operator $2S_n$.

$$|\mu_n|^2(y) = \exp[-(S_n y, y)].$$

From Theorem 3.3 it follows that there exists a compact sequence U_n of S-operators such that

$$1 - |\mu_n|^2(y) \leqslant (U_n y, y) + \epsilon.$$

Hence there is a δ such that for any n, $(U_n y, y) \leqslant \delta$ implies $(S_n y, y) \leqslant 1$. From this we can deduce that

$$(S_n y, y) \leqslant \delta^{-1}(U_n y, y).$$

Since δ is independent of n and U_n is compact, S_n is also compact.

Lemma 6.2 : *Let μ be a symmetric infinitely divisible distribution with $\mu = [0, 0, M]$. Further, let M be concentrated in the unit sphere. Then*

$$\int \|x\|^4 \, d\mu \leqslant \int \|x\|^4 \, dM + 3[\int \|x\|^2 \, dM]^2 < \infty.$$

Proof : It is enough to prove the theorem when M is finite since the other case can be obtained by limit transition. Let $M(x) = t$ and F be the distribution such that $F(X) = 1$ and $M = tF$. Then

$$\mu = e^{-t} \sum \frac{t^r F^r}{r!}$$

$$\int \|x\|^4 \, dF^r = E \, \|X_1 + \ldots X_r\|^4$$
$$= r \, E \, \|X_1\|^4 + r(r-1)[E \, \|X_1\|^2]^2 + 2r(r-1) \, E(X_1, X_2)^2$$
$$= rE \, \|X_1\|^4 + 3r(r-1)(E \, \|X_1\|^2)^2.$$

Expectation is with respect to F and $X_1, \ldots X_r$ are independent random variables in X with the same distribution F. Terms with zero expectation have been omitted.

$$\int \|x\|^4 \, d\mu \leqslant e^{-t} \sum \frac{rt^r \, E \, \|X_1\|^4}{r!} + 3 \, e^{-t} \sum \frac{r(r-1)t^r(E \, \|X_1\|^2)^2}{r!}$$
$$= tE \, \|X_1\|^4 + 3t^2(E \, \|X_1\|^2)^2$$
$$= t \int \|X\|^4 \, dF + 3t^2[\int \|X\|^2 \, dF]^2$$
$$= \int \|X\|^4 \, dM + 3[\int \|X\|^2 \, dM]^2.$$

Lemma 6.3 : *Let μ_n be symmetric infinitely divisible distributions such that*

$$\mu_n = [0, 0, M_n]$$

with M_n vanishing outside the sphere $\|X\| \leqslant 1$ for all n. If μ_n is compact then

$$\sup_n \int \|X\|^4 \, d\mu_n < \infty.$$

Proof: Since M_n is concentrated in the unit sphere

$$\int \|X\|^4 \, dM_n \leqslant \int \|X\|^2 \, dM_n.$$

Theorem 5.6 implies that $\quad \sup_n \int \|X\|^2 \, dM_n < \infty$

and hence an application of Lemma 6.2 will complete the proof.

Remark 6.1 : In the same manner as Lemma 6.2 it can be shown that if M is symmetric and vanishes outside the sphere $\|X\| \leqslant 1$ for

$$\mu = [0, 0, M].$$

We have $\quad\quad \int (x, y)^2 d\mu = \int (x, y)^2 dM$ for all y.

Lemma 6.4 : Let μ_n be a weakly conditionally compact sequence of symmetric distribution such that

$$\sup_n \int \|X\|^4 \, d\mu_n < \infty.$$

Then if S_n is the covariance operator of μ_n, $\{S_n\}$ is compact.

Proof: If θ_n is a sequence of distributions on the real line such that $\theta_n \Longrightarrow \theta$ and $\int x^2 \, d\theta_n$ is uniformly bounded then

$$\int x d\theta_n \to \int x d\theta \text{ as } n \to \infty.$$

Theorem 3.5 can be applied now and the lemma follows at once.

Theorem 6.1 : Let μ_n be symmetric distributions that are infinitely divisible with representations

$$\mu_n = [0, S_n, M_n].$$

Then in order that μ_n be compact it is necessary and sufficient that

(i) M_n restricted outside any neighbourhood of the identity is weakly conditionally compact,

(ii) T_n as defined in (6.1) is compact.

Proof: We will first prove sufficiency. Let unit sphere be chosen as the neighbourhood and let us write $M_n = M_n^{(1)} + M_n^{(2)}$ where $M_n^{(1)}$ and $M_n^{(2)}$ are respectively the restrictions of M_n inside the unit sphere and outside the unit sphere. Since $M_n^{(2)}$ is weakly conditionally compact and $F_n \Longrightarrow F$ implies $e(F_n) \Longrightarrow e(F)$ it is enough to show that the distributions

$$\lambda_n = [0, S_n, M_n^{(1)}]$$

form a compact sequence $\int (x, y)^2 d\lambda_n(x) = 2(S_n y, y) + \int (x, y)^2 \, dM_n^{(1)}(x)$

$$= 2(S_n y, y) + \int_{\|x\| \leqslant 1} (x, y)^2 \, dM_n(x)$$

$$= (T_n y, y).$$

Since T_n is compact sufficiency follows from Theorem 3.4. Necessity of (i) is a consequence of Theorem 5.3 and (ii) follows from Lemmas 6.1, 6.2 and 6.4.

Theorem 6.2 : *In order that a sequence μ_n of infinitely divisible distributions with representations*

$$\mu_n = [x_n, S_n, M_n]$$

be shift compact it is necessary and sufficient that

(i) *M_n restricted to the complement of any neighbourhood N of the origin is weakly conditionally compact,*

(ii) *T_n as defined in (6.1) is compact.*

Proof : Since μ_n is shift compact if and only $|\mu_n|^2$ is compact what we need are conditions for the compactness of

$$|\mu_n|^2 = [0, 2S_n, M_n + \bar{M}_n].$$

In addition we have $\displaystyle\int_{\|x\| < 1} (x, y)^2 \, d(M_n + \bar{M}_n) = 2 \int_{\|x\| < 1} (x, y)^2 dM_n.$

Hence the theorem follows from 6.1.

Theorem 6.3 : *In order that μ_n with the representations*

$$\mu_n = [x_n, S_n, M_n]$$

be weakly conditionally compact it is necessary and sufficient that in addition to the conditions of Theorem 6.2 x_n should be compact in X.

Proof : In order to prove the theorem it suffices to show that whenever

$$\mu_n = [x_n, S_n, M_n]$$

is shift compact, $\lambda_n = [0, S_n, M_n]$

is compact. Let $F_n \Longrightarrow F$. Then the following convergence takes place in norm.

$$\int \frac{x}{1 + \|x\|^2} \, dF_n \to \int \frac{x}{1 + \|x\|^2} \, dF.$$

Hence we can assume that M_n vanishes for all n outside the sphere $\|x\| \leqslant 1$. Let us now consider

$$f_n(y) = \int K(x, y) \, dM_n(x)$$

$$|f_n(y)| \leqslant \tfrac{1}{2} \int_{\|x\| < 1} (x, y)^2 \, dM_n(x) + \int_{\|x\| > 1} \frac{|(x, y)| \, \|x\|^2}{1 + \|x\|^2} \, dM_n(x)$$

$$\leqslant \tfrac{1}{2}(T_n y, y) + \int |(x, y)| \, \|x\|^2 \, dM_n(x)$$

$$\leqslant \tfrac{1}{2}(T_n y, y) + [\int (x, y)^2 \|x\|^2 \, dM_n \int \|x\|^2 \, dM_n]^{\frac{1}{2}}$$

$$\leqslant \tfrac{1}{2}(T_n y, y) + C(T_n y, y)^{\frac{1}{2}}.$$

Hence given any $\rho > 0$ there is a $\delta > 0$ such that $|f_n(y)| \leqslant \rho$ whenever $(T_n y, y) < \delta$. But since $\lambda_n(y) = \exp[f_n(y)]$, for any $\epsilon > 0$ there exists a $\rho > 0$ such that $1 - R\lambda_n(y) \leqslant \epsilon$ whenever $|f_n(y)| \leqslant \rho$. Consequently, for any $\epsilon > 0$ there exists a $\delta > 0$ such that whenever $(T_n y, y) \leqslant \delta$ it follows that $1 - R\lambda_n(y) \leqslant \epsilon$. Here R denotes the real part. Now Lemma 5.1 and Theorem 3.5 imply the validity of the theorem since $\{T_n\}$ is compact.

Remark 6.2 : In defining the operator T we could have taken any bounded sphere around the origin instead of the unit sphere. When M_n restricted outside any neighbourhood of the origin is known to be weakly conditionally compact, the compactness of T_n when it is based on some sphere implies the compactness of T_n when it is based on any finite sphere.

Since the representation is unique one can give conditions for the convergence of

$$\mu_n = [x_n, S_n, M_n]$$

to the distribution $\mu = [x, S, M]$

in terms of $[x_n, S_n, M_n]$ and $[x, S, M]$. However, we will mention only the following.

Theorem 6.4 : *Let μ_n be a sequence such that μ_n has the representation*

$$\mu = [x_n, S_n, M_n].$$

If $\mu_n \Longrightarrow \mu$, μ is Gaussian if and only if

$$M_n(N') \to 0 \text{ as } n \to \infty$$

for every neighbourhood N of the origin.

Proof : Let μ be Gaussian. Since μ cannot be written as $e(F) * \lambda$ with λ infinitely divisible it follows that $M_n(N') \to 0$ as $n \to \infty$ for every neighbourhood N of the origin, conversely, if $M_n(N') \to 0$ for every neighbourhood N in exactly the same manner as in the proof of Theorem 6.1 of Parthasarathy *et al* (1962a) it can be shown that $\mu(y)$ is of the form

$$\mu(y) = \exp [i (x_0, y) - (Sy, y)]$$

which shows that μ is Gaussian.

7. ACCOMPANYING LAWS

Definition 7.1 : A sequence α_{nj} of distributions $j = 1, 2, ..., k_n$, $n = 1, 2, ...$ is said to be uniformly infinitesimal if for any neighbourhood N of the origin

$$\underset{n \to \infty}{\text{lt}} \underset{1 \leqslant j \leqslant k_n}{\inf} \alpha_{nj}(N) = 1.$$

Theorem 7.1 : *In order that α_{nj} be uniformly infinitesimal it is necessary that*

$$\underset{n \to \infty}{\text{lt}} \underset{1 \leqslant j \leqslant k_n}{\sup} \underset{\|y\| \leqslant K}{\sup} |\alpha_{nj}(y) - 1| = 0$$

for every constant K.

Proof : This is immediate from Theorem 5.4.

Theorem 7.2 : *Let α_{nj} be uniformly infinitesimal symmetric distributions with non-negative characteristic function. Let*

$$\mu_n = \prod_{j=1}^{k_n} \alpha_{nj}$$

and λ_n be defined as

$$\lambda_n = \prod_{j=1}^{k_n} e(\alpha_{nj}).$$

In order that $\mu_n \Longrightarrow \mu$ it is necessary and sufficient that $\lambda_n \Longrightarrow \mu$.

Proof : Let μ_n be compact. Since the inequality $e^{x-1} \geqslant x$ is valid for all x we have

$$e(\alpha_{nj})(y) \geqslant \alpha_{nj}(y) \quad \text{for } j = 1, 2, ..., k_n, \quad n = 1, 2 ...$$

since $\alpha_{nj}(y)$ is non-negative

$$\lambda_n(y) \geqslant \mu_n(y).$$

Or

$$1 - \lambda_n(y) \leqslant 1 - \mu_n(y).$$

From the compactness of μ_n and Theorem 3.3 it follows that λ_n is compact. Now let λ_n be compact. It follows from Theorem 6.2 and the remark made after Theorem 6.3 that

    (i) F_n restricted to N' is weakly conditionally compact for every neighbourhood N;

    (ii) the sequence S_n of operators defined by

$$(S_n y, y) = \int\limits_{\|x\| \leqslant t} (x, y)^2 dF_n(x)$$

is compact for every t.

Here F_n denotes the sum $\alpha_{n1} + \alpha_{n2} + ... \alpha_{nk_n}$. We will now show that μ_n is compact. We have

$$1 - \mu_n(y) \leqslant \sum_{j=1}^{k_n} [1 - \alpha_{nj}(y)] = \int [1 - \cos (x, y)] dF_n(x)$$

$$= \int\limits_{\|x\| < t} [1 - \cos (x, y)] dF_n(x) + \int\limits_{\|x\| > t} [1 - \cos(x, y)] dF_n(x)$$

$$= \tfrac{1}{2} \int\limits_{\|x\| < t} (x, y)^2 dF_n(x) + 2F_n[\|x\| > t] = \tfrac{1}{2}(S_n y, y) + 2F_n[\|x\| > t].$$

Since F_n is weakly compact outside any neighbourhood we can choose t such that for all n

$$F_n[\|x\| > t] < \epsilon/2.$$

Since for that fixed t, S_n is compact Theorem 3.3 shows that μ_n is compact.

    We will now complete the proof by showing that whenever λ_n is compact, for every costant k

$$\sup_{\|y\| \leqslant K} |\lambda_n(y) - \mu_n(y)| \to 0.$$

To this end it is enough to show that

$$\sup_{\|y\| \leqslant K} \sup_n \sum_{j=1}^{k_n} [1 - \alpha_{nj}(y)] < \infty.$$

But the expression is equal to

$$\sup_{\|y\| \leqslant K} \sup_n \int [1 - \cos (x, y)] dF_n \leqslant \sup_{\|y\| \leqslant K} \sup_n \int\limits_{\|x\| \leqslant 1} [1 - \cos (x, y)] dF_n + \sup_n 2F_n(\|x\| > 1)$$

$$\leqslant \tfrac{1}{2} K^2 \sup_n \int\limits_{\|x\| \leqslant 1} \|x\|^2 dF_n + 2 \sup_n F_n(\|x\| > 1)$$

$$< \infty.$$

The last step follows since $\lambda_n = e(F_n)$ is compact.

Lemma 7.1 : *Let α_{nj} be uniformly infinitesimal. Then if x_{nj} is defined by the relation*

$$x_{nj} = \int_{\|x\| \leqslant 1} x \, d\alpha_{nj}$$

then

$$\sup_{1 \leqslant j \leqslant k_n} \|x_{nj}\| \to 0 \text{ as } n \to \infty.$$

Proof : Let ϵ be arbitrary and V the sphere $\|x\| \leqslant \epsilon$. Then

$$\|x_{nj}\| \leqslant \|\int_V d\alpha_{nj}\| + \|\int_{\epsilon < \|x\| \leqslant 1} x \, d\alpha_{nj}\| \leqslant \epsilon + \alpha_{nj}(V').$$

Hence

$$\limsup_{n \to \infty} \sup_{1 \leqslant j \leqslant k_n} \|x_{nj}\| \leqslant \epsilon.$$

Since ϵ is arbitrary the lemma follows.

Lemma 7.2 : *Let α_{nj} be uniformly infinitesimal. Let x_{nj} be defined as in Lemma 4.6. Then if $\theta_{nj} = \alpha_{nj} * (-x_{nj})$ there exists a n_0 such that for all $1 \leqslant j \leqslant k_n$ and $n \geqslant n_0$ we have*

$$\|\int_{\|x\| \leqslant 1} x \, d\theta_{nj}\| \leqslant 2\theta_{nj}[\|x_{nj}\| > 1].$$

Proof : Let n_0 be so chosen that

$$\sup_{1 \leqslant j \leqslant k_n} \|x_{nj}\| \leqslant 1/4 \text{ for all } n > n_0.$$

Then

$$\int_{\|x\| \leqslant 1} x \, d\theta_{nj} = \int_{\|x - x_{nj}\| \leqslant 1} (x - x_{nj}) d\alpha_{nj}$$

$$= \int_{\|x - x_{nj}\| \leqslant 1} x \, d\alpha_{nj} - \int_{\|x\| \leqslant 1} x \, d\alpha_{nj} + x_{nj}\alpha_{nj}[\|x - x_{nj}\| > 1].$$

Therefore, for $n > n_0$ and $1 \leqslant j \leqslant k_n$,

$$\|\int_{\|x\| \leqslant 1} x \, d\theta_{nj}\| \leqslant \int_{[\|x - x_{nj}\| \leqslant 1 \wedge \|x\| \leqslant 1]} \|x\| d\alpha_{nj} + \|x_{nj}\| \theta_{nj}[\|x\| > 1]$$

$$\leqslant \int_{\frac{3}{4} < \|x\| < \frac{5}{4}} \|x\| d\alpha_{nj} + \tfrac{1}{4} \theta_{nj}[\|x\| > 1]$$

$$\leqslant \tfrac{5}{4} \alpha_{nj}[\|x\| > 3/4] + \tfrac{1}{4} \theta_{nj}[\|x\| > 1]$$

$$\leqslant \tfrac{5}{4} \theta_{nj}[\|x\| > 1] + \tfrac{1}{4} \theta_{nj}[\|x\| > 1]$$

$$\leqslant 2\theta_{nj}[\|x\| > 1].$$

Lemma 7.3 : *Let F_n be a sequence of σ-finite measures such that F_n restricted to N' is finite and weakly conditionally compact for every neighbourhood N of the origin. Then for any $\epsilon > 0$ there exists a compact set K such that*

$$F_n(K') \leqslant \epsilon \text{ for } n = 1, 2, \ldots$$

Proof : Let ϵ be any positive number. Choose a sequence N_r of neighbourhoods of the origin decreasing to the origin. Let A_r be defined as $N_{r-1} - N_r$, N_0 being taken as the whole space. From the conditions of the lemma it is possible to find a compact set K_r in A_r such that for all n,

$$F_n(A_r - K_r) \leqslant \frac{\epsilon}{2^r}.$$

Let K be defined as
$$K = \bigcup_{r=1}^{\infty} K_r \cup \{0\}.$$

Since
$$K \cap N'_r = K_1 \cup K_2 \ldots \cup K_r$$

and N_r decreases to the origin it follows that K is compact and

$$F_n(K') \leqslant \sum_{r=1}^{\infty} F_n(A_r - K_r) \leqslant \epsilon.$$

We now proceed to prove the main theorem of the section.

Let α_{nj} be uniformly infinitesimal sequence of distributions on X. Let μ_n, x_n, θ_{nj}, λ_n be defined as follows.

$$\mu_n = \prod_{j=1}^{k_n} \alpha_{nj}$$

$$x_{nj} = \int_{\|x\| < 1} x \, d\alpha_{nj}$$

$$\theta_{nj} = \alpha_{nj}{}^*(-x_{nj})$$

$$\lambda_n = \prod_{j=1}^{k_n} e(\theta_{nj})^* \left(\sum_{j=1}^{k_n} x_{nj} \right)$$

In what follows we will adopt the above notation.

Theorem 7.3 : *If μ_n is shift compact so is λ_n.*

Proof : Since μ_n is shift compact $|\mu_n|^2$ is compact. But

$$|\mu_n|^2 = \prod_{j=1}^{k_n} |\alpha_{nj}|^2.$$

It follows now from Theorem 6.2 that if one defines

$$F_n = \sum_{j=1}^{k_n} |\alpha_{nj}|^2 = \sum_{j=1}^{k_n} |\theta_{nj}|^2. \qquad \ldots \ (7.1)$$

Then $e(F_n)$ is compact. We can now apply Theorem 6.2 and Lemma 7.3 and deduce that for any $\epsilon > 0$ there exists a compact set K such that

$$F_n(K') \leqslant \epsilon \quad \text{for } n = 1, 2, \ldots \qquad \ldots \ (7.2)$$

Let us now define the S-operators T_n by the formula

$$(T_n y, y) = \int_{\|x\| < t} (x, y)^2 \, dF_n(x).$$

Then for any finite t the sequence T_n is compact.
Let us now define G_n as

$$G_n = \sum_{j=1}^{k_n} \theta_{nj}. \qquad \ldots \ (7.3)$$

In order to complete the theorem we have to show that $e(G_n)$ is shift compact or

    (a) G_n is weakly conditionally compact when restricted outside any neighbourhood of the identity;

    (b)   if the S-operators S_n are defined as

$$(S_n y, y) = \int_{\|x\| < 1} (x, y)^2 dG_n$$

then S_n is compact.

Since α_{nj} is uniformly infinitesimal by Lemma 7.1 θ_{nj} are also uniformly infinitesimal. Hence for any $\epsilon > 0$ there exists a compact set C such that

$$\theta_{nj}(C) \geqslant 1-\epsilon \text{ for all } n \text{ and } 1 \leqslant j \leqslant k_n.$$

In the same manner as in the proof of Theorem 5.1 of Parthasarathy *et al* (1962a) we have for all n and $1 \leqslant j \leqslant k_n$

$$|\theta_{nj}|^2(K') = \int \theta_{nj}(K'+x)d\theta_{nj}(x) \geqslant \int_C \theta_{nj}(K'+x)d_{nj}(x)$$

$$\geqslant (1-\epsilon) \inf_{x \in C} \theta_{nj} (K'+x) = (1-\epsilon)[1-\sup_{x \in C} \theta_{nj} (K+x)]$$

$$= (1-\epsilon)\theta_{nj}(\overline{K+C'}) = (1-\epsilon) \theta_{nj}(K_1'). \qquad \qquad \ldots \quad (7.4)$$

where K_1 is another compact set.

In a similar manner it can be shown that if V and N are two neighbourhoods of the origin such that $V+V \subseteq N$ and

$$\theta_{nj}(V) \geqslant 1-\epsilon \text{ for all } n \geqslant n_0, 1 \leqslant j \leqslant k_n \qquad \ldots \quad (7.5)$$

then for $\qquad n \geqslant n_0 \text{ and } 1 \leqslant j \leqslant k_n \; |\theta_{nj}|^2(V') \geqslant (1-\epsilon)\theta_{nj} (N') \qquad \ldots \quad (7.6)$

(7.2), (7.3) and (7.4) imply that for any $\epsilon > 0$ there exists a compact set K_1 such that

$$G_n(K_1') \leqslant \epsilon \text{ for all } n. \qquad \qquad \ldots \quad (7.7)$$

On the other hand (7.1), (7.3), (7.6) and the weak compactness of F_n when restricted to the complement of any neighbourhood of the origin imply that for any neighbourhood N of the origin,

$$\sup_n G_n(N') < \infty. \qquad \qquad \ldots \quad (7.8)$$

(7.7) and (7.8) prove that G_n is weakly conditionally compact when restricted outside any neighbourhood of the origin. To prove (b) let us consider

$$\int_{\|x\| \leqslant 2} (x, y)^2 d |\theta_{nj}|^2(x) = \iint_{\|x_1-x_2\| \leqslant 2} (x_1-x_2, y)^2 d\theta_{nj}(x_1)d\theta_{nj}(x_2)$$

$$\geqslant \iint_{\|x_1\| < 1, \|x_2\| < 1} (x_1-x_2, y)^2 d\theta_{nj}(x_1)d\theta_{nj}(x_2)$$

$$= 2\theta_{nj}(\|x\| \leqslant 1) \int_{\|x\| < 1} (x, y)^2 d\theta_{nj}(x) - 2[\int_{\|x\| < 1} (x, y)d\theta_{nj}]^2.$$

Since θ_{nj} is uniformly infinitesimal we can assume that $\theta_{nj}(\|x\| \leqslant 1) \geqslant \frac{1}{2}$ for all suitably large n and for all $1 \leqslant j \leqslant k_n$. Hence

$$(S_n y, y) \leqslant (T_n y, y) + 2(U_n y, y)$$

where

$$(U_n y, y) = \sum_{j=1}^{k_n} [\int \cdot (x, y) d\theta_{nj}(x)]^2.$$

Since we know that T_n is compact in order to show that S_n is compact it is enough to prove that U_n is compact. We will show now that trace of U_n tends to zero as $n \to \infty$, and this would complete the proof.

Let us put

$$y_{nj} = \int_{\|x\| < 1} x d\theta_{nj}$$

Then for the trace of U_n we have

trace of

$$U_n = \sum_{j=1}^{k_n} \|y_{nj}\|^2 \leqslant (\sup_{1 < j \leqslant k_n} \|y_{nj}\|) \left(\sum_{j=1}^{k_n} |y_{nj}| \right).$$

From Lemmas 7.1, 7.2 and 7.8 it follows that trace $U_n \to 0$ as $n \to \infty$.

Theorem 7.4 : *If λ_n is shift compact so is μ_n.*

Proof : $\quad 1 - |\mu_n|^2(y) = 1 - \prod_{j=1}^{k_n} |\theta_{nj}|^2(y) \leqslant \sum_{j=1}^{k_n} [1 - |\theta_{nj}|^2(y)]$

$$\leqslant 2 \sum_{j=1}^{k_n} [1 - \text{Real } \theta_{nj}(y)] = 2 \sum_{j=1}^{k_n} \int [1 - \cos(x, y)] d\theta_{nj}$$

$$= 2 \int [1 - \cos(x, y)] dG_n(x)$$

$$= 2 \int_{\|x\| < t} [1 - \cos(x, y)] dG_n(x) + 2G_n[\|x\| > t]$$

$$\leqslant \int_{\|x\| < t} (x, y)^2 dG_n(x) + 2G_n[\|x\| > t]$$

$$= (S_n y, y) + 2G_n[\|x\| > t]. \qquad \qquad \dots \text{(7.9)}$$

From the shift compactness of λ_n for any $\epsilon > 0$ we can choose t such that

$$G_n[\|x\| > t] \leqslant \frac{\epsilon}{2} \text{ for all } n \qquad \qquad \dots \text{(7.10)}$$

and the sequence of operators S_n defined by

$$(S_n y, y) = \int_{\|x\| < t} (x, y)^2 dG_n(x)$$

is compact. From (7.9), and (7.10) the compactness of S_n and Theorem 3.3 the present theorem follows.

Theorem 7.5 : *Let λ_n be shift compact. Then for any finite number t we have*

$$\underset{n \to \infty}{\text{lt}} \sup_{\|y\| \leqslant t} |\lambda_n(y) - \mu_n(y)| = 0.$$

Proof : Since θ_{nj} is uniformly infinitesimal it follows from Theorem 7.1 that

$$\underset{n \to \infty}{\text{lt}} \sup_{\|y\| \leqslant t} \sup_{1 < j \leqslant k_n} |\theta_{nj}(y) - 1| = 0.$$

Hence it is enough to show that $\displaystyle\sup_{n} \sup_{\|y\| \leqslant t} \sum_{j=1}^{k_n} |\theta_{nj}(y) - 1| < \infty.$ ... (7.11)

To this end we have

$$\theta_{nj}(y) - 1 = \int [e^{i(x,y)} - 1] d\theta_{nj}$$

$$= \int_{\|x\| \leqslant 1} [e^{i(x,y)} - 1 - i(x, y)] d\theta_{nj} + i \int_{\|x\| \leqslant 1} (x, y) d\theta_{nj} + \int_{\|x\| > 1} [e^{i(x,y)} - 1] d\theta_{nj}$$

$$|\theta_{nj}(y) - 1| \leqslant \tfrac{1}{2} \int_{\|x\| \leqslant 1} (x, y)^2 d\theta_{nj} + \|y\| \| \int_{\|x\| \leqslant 1} x d_{nj} \| + 2\theta_{nj}[\|x\| > 1].$$

(7.11) follows at once from the compactness of S_n, Lemma 7.2 and (7.8).

Theorem 7.6 : *In order that $\mu_n * x_n \Longrightarrow \mu$, where x_n are arbitrary points in X, it is necessary and sufficient that $\lambda_n * x_n \Longrightarrow \mu$.*

Proof : This is an immediate consequence of Theorems 7.3, 7.4, 7.5 and 5.5.

Theorem 7.7 : *Limit distribution of sums of independent uniformly infinitesimal random variables in X is infinitely divisible.*

Proof : Theorems 7.6 and 5.1 imply the present theorem since λ_n is infinitely divisible for each n.

Theorem 7.8 : *Let $\mu_n * x_n \Longrightarrow \mu$. In order that μ may be Gaussian it is necessary and sufficient that for each neighbourhood N*

$$\underset{n \to \infty}{\text{lt}} \ G_n(N') = 0,$$

where

$$G_n = \sum_{j=1}^{k_n} \theta_{nj}.$$

Proof : This follows immediately from Theorems 7.6 and 6.4.

References

Bochner, S. (1955) : *Harmonic Analysis in Probability*, University of California Press.

————— (1958) : General analytical setting for the central limit theory of probability. Golden Jubilee Commemoration Volume of the Calcutta Mathematical Society, Part I, pp. 111-128.

Gnedenko, B. V. and Kolmogorov, A. N. (1954) : *Limit Distributions for Sums of Independent Random Variables*, Addison-Wesley, Cambridge.

Halmos, P. R. (1950) : *Measure Theory*, Van Nostrand, Princeton.

Hunt, G. A. (1956) : Semigroups of measures on Lie groups. *Trans. Amer. Math. Soc.*, **81**, 264-293.

Kloss, B. M. (1961) : Limiting distributions on bicompact abelian groups. *Teorya Veroyot i Premen.*, **6**, 392-421.

Levy, P. (1939) : L'addition des variables aleatoires des variables definies sur une cerconforence. *Bull. Soc. Math.*, **67**, 1-41.

————— (1954) : *Theorie de L'addition des Variables Aleatoires*, Gauthier—Villars.

Parthasarathy, K. R., Rangarao, R. and Varadhan, S. R. S. (1962) : On the category of indecomposable distributions on topological groups. *Trans. Amer. Math. Soc.*, **102**, No. 2, 200-217.

————— (1962a) : On the semigroup of distributions on topological groups. To appear in *Illinois J. Math.*

Prohorov, Yu. V. (1956) : Convergence of random processes and limit theorems in the theory of probability. *Teorya Veroyot. i. Premen.*, **1**, 177-238.

Rangarao, R. (1960) : Some problems in probability theory. Thesis submitted to the Calcutta University (unpublished).

Sazanov, V. (1958) : On characteristic functionals. *Teorya Veroyot i. Premen.*, **3**, 201-205.

Takano, K. (1955) : Central convergence criterion in the multidimensional case. *Ann. Inst. Stat. Math.*, **7** 95-102.

Varadarajan, V. S. (1958) : Weak convergence of measures on separable metric spaces. *Sankhyā*, **19**, 15-22.

Vorobev, N. N. (1954) : Addition of independent random variables on finite abelian groups. *Mat. Sbornik*, **34**, 89-216.

ON THE USE OF A DISTRIBUTION-FREE PROPERTY IN DETERMINING A TRANSFORMATION OF ONE VARIATE SUCH THAT IT WILL EXCEED ANOTHER WITH A GIVEN PROBABILITY*

By SAM C. SAUNDERS

Boeing Scientific Research Laboratories, Seattle, Washington

SUMMARY. Let X and Y be independent random variables, with continuous distributions $F \varepsilon \mathscr{F}$ and $G \varepsilon \mathscr{G}$, on the sample space $\mathscr{X}$. Let ω be a homeomorphism from $\mathscr{X}$ onto itself and define

$$H(\omega) = \int F(\omega) dG.$$

From samples of X and Y we form $\tilde{F}$ and $\tilde{G}$ which are estimates of F and G, respectively, and define an estimate of H, say $\tilde{H}$,

$$\tilde{H}(\omega) = \int \tilde{F}(\omega) d\tilde{G}$$

for $\omega \varepsilon \Omega$, a class of homeomorphisms linearly ordered by H.

Interpreting $\omega(Y)$ as the strain under use ω and X as the strength (of some device), then $H(\omega) = P[X < \omega(Y)]$ represents the unreliability and $\tilde{H}(\omega)$ is an estimate of it. If we use $\tilde{H}$ to determine a use $\tilde{\omega}$, then what is the probability that the true unreliability $\tilde{H}(\omega)$ is too large? We examine this problem under the assumptions that $\tilde{F}(F^{-1})$ and $\tilde{G}(G^{-1})$ are distribution-free with respect to $\mathscr{F}$ and $\mathscr{G}$ respectively. This provides an answer in some cases and allows one to obtain stochastic bounds in others.

1. GENESIS OF THE PROBLEM

We were asked to consider the problem of determining the unreliability of a missile fuel tank for a given weight of liquid hydrocarbon fuel to be installed. The volume of the tank, the specific weight of a batch of fuel and the temperature environment were stochastic variables.

In the temperature-pressure range for this specific problem considerations of the physical and chemical properties of the fuel show that the design pressure is exceeded if the following event obtains : $[X < wg(Y)]$, where X is the random volume of the tank, w is the weight of the fuel installed and g is a known function (determined by the bulk modulus of the fuel, the temperature variation of range and the design pressure of the tank) of Y the random specific weight of the fuel batch. If a weight w of fuel is installed, the design unreliability of the tank, i.e., the probability of the pressure exceeding the design specification, is $H(w) = P[X < wg(Y)]$. If the distribution of X was F and that of Y was G, both known, then we could express

$$H(w) = \int_0^{\infty} F(wg(x)) dG(x).$$

239

5

This equation defines the function H and hence

(a) for a given weight w of fuel installed we can determine the design unreliability $H(w)$,

(b) for a specified design unreliability of at most ϵ we can seek the maximum weight w for which we have $H(w) \leqslant \epsilon$.

However, in practice the distributions F and G are usually not known and we have only sample values of tank volumes and the sample values of the specific weights of fuel batches with which to arrive somehow at answers which correspond to the cases (a) and (b).

It is the study of the statistical problems for the situations (a) and (b), within the more inclusive formulation of the problem that we propose, which constitutes the subject matter of this note.

2. Introduction and related results

Let $(\mathscr{X}, \mathfrak{a})$ be a measurable space for which $(\mathscr{X}, <)$ is a partially ordered set, i.e., the relation is reflexive, transitive and such that $x < y$ and $y < x$ imply $x = y$. If the relation is measurable, i.e., for all $x \epsilon \mathscr{X}$ the set $\{y \epsilon \mathscr{X} : y < x\}$ is in $\mathfrak{a}$, then a distribution can be defined on $\mathscr{X}$ for each measure P on $\mathfrak{a}$ by $F(x) = P[X < x]$, with the usual interpretation of X as the identity random variable (r.v.).

If for each $x \epsilon \mathscr{X}$ the set $[X = x]$ has measure zero, then the distribution F is continuous. Now $\mathscr{X}$ is called here a positive sample space if and only if (iff) for each $x < y$, $x \neq y$ we have the set $[X > x] \bigcap [X < y]$ has positive P-measure for all probability measures P under consideration.

Let X and Y be random variables taking values in the positive sample space $\mathscr{X}$ with continuous distributions F and G respectively. (The generalization to Y taking values in a space different from $\mathscr{X}$ will be seen to be immediate.) Let Ω be a set of transformations on $\mathscr{X}$ onto itself. For each $\omega \epsilon \Omega$ we have the probability that X precedes $\omega(Y)$ according to $<$ given by

$$H(\omega) = P[X < \omega(Y)] = \int F\omega dG \qquad \dots \quad (2.1)$$

where we make the convention that juxtaposition of functions refers to composition and integrals are understood to be over the entire space $\mathscr{X}$.

Let us assume that while F and G are unknown we can obtain samples of X and Y from which we form, respectively, the estimates $\tilde{F}$ and $\tilde{G}$ of the distributions. Thus we define an estimate of H, say $\tilde{H}$, by

$$\tilde{H}(\omega) = \int \tilde{F}\omega d\tilde{G} \qquad \dots \quad (2.2)$$

for each $\omega \epsilon \Omega$ for which the integrals exist.

Some questions of interest are : In what sense is $\tilde{H}$ a good estimate of H ? If ω is known, what is the distribution of $\tilde{H}(\omega)$? But primarily we want the maximum

transformation $\tilde{\omega}$ (as a function of $\tilde{H}$) for which for each $F \epsilon \mathcal{F}$, $G \epsilon \mathcal{G}$ we have

$$P[H(\tilde{\omega}) > \epsilon] \leqslant \alpha$$

where α and ϵ are small and specified in advance. That is, we are interested in the problem of using $\tilde{H}$ to obtain what would be tolerance limits were H a probability distribution on Ω.

If Ω consists of a single point ω, we can without loss of generality assume it is the identity transformation. We then have the problem of estimating $p = P[X < Y]$ from samples of X and Y.

Suppose that $\mathcal{X}$ is the real line with the usual ordering and $\mathcal{F} = \mathcal{G}$ is the class of continuous distributions. Then it is well known that the empiric distribution, say $\hat{G}$, is the unique minimum variance unbiased estimate of G. But, further, $\hat{p}$ defined analogously to (2.2) is the unique minimum variance unbiased estimate of p and $mn\hat{p}$ is the Mann-Whitney statistic where m and n are the sample sizes used in computing $\hat{F}$ and $\hat{G}$.

The problem analogous to ours, i.e., involving $\hat{p}$ and p, has been studied and bounds for the corresponding estimates have been obtained for large sample sizes under the assumption that one of the distributions is known (Birnbaum, 1956) and assuming that neither of the distributions is known (Birnbaum and McCarty, 1958).

3. Results

Let $\leqslant$ be a partial ordering on the set of all transformations of $\mathcal{X}$ onto itself for which $(\Omega, \leqslant)$ is a linearly ordered subset which is order complete. That is, $(\Omega, \leqslant)$ is a partially ordered set for which any two elements are comparable and each non void subset of Ω which has a lower bound has an infimum.

For a sample of independent r.v.'s each with the same distribution $F \epsilon \mathcal{F}$, the estimate $\tilde{F}$ (which is a measurable function of the sample) of F is *ample for $\mathcal{F}$* iff the random function $\tilde{F}F^{-1}$ has the same probability law for every $F \epsilon \mathcal{F}$.

We now make our assumptions :

(1) $\mathcal{F}$ and $\mathcal{G}$ are classes of continuous distributions on the partially ordered positive sample space $(\mathcal{X}, <)$.

(2) $(\Omega, \leqslant)$ is a linearly ordered complete space of transformations on $\mathcal{X}$ such that

$$\omega_1 \leqslant \omega_2 \text{ implies } \omega_1(x) < \omega_2(x) \text{ for all } x \epsilon \mathcal{X}. \qquad \dots \text{ (3.1)}$$

(3) $\tilde{F}$ and $\tilde{G}$ are ample estimates for $F \epsilon \mathcal{F}$ and $G \epsilon \mathcal{G}$, respectively.

Since there always exists a set of transformations on $\mathcal{X}$, say Γ, such that $\mathcal{F} = (F_0\gamma : \gamma \epsilon \Gamma)$ and similarly $\mathcal{G} = \{G_0\lambda : \lambda \epsilon \Lambda\}$ where Γ and Λ each contain the identity function, it is seen in this representation that $\tilde{F}$ is ample iff $\tilde{F}\gamma^{-1}$ (or equivalently iff $\tilde{\gamma}\gamma^{-1}$) is distribution-free with respect to $\mathcal{F}$. We shall adopt this representation wherever convenient, without comment, in what follows.

The set

$$\Phi = \{\gamma\omega\lambda^{-1} : \lambda\epsilon\Lambda, \ \gamma\epsilon\Gamma, \ \omega\epsilon\Omega\} \qquad \ldots \ (3.2)$$

is a set of transformations on $\mathscr{X}$ which is, in general, partially ordered by $\leqslant$. Let

$$H_0(\phi) = \int F_0\phi dG_0, \qquad \ldots \ (3.3)$$

$$\tilde{H}_0(\phi) = \int \tilde{F}\gamma^{-1}\phi d\tilde{G}\lambda^{-1} \qquad \ldots \ (3.4)$$

be two functions, the second random, defined for each $\phi \epsilon \Phi$ for which the integrals exist. (The functions $\tilde{F}$ and $\tilde{G}$ may both contain discontinuities and these cannot be made to coincide.)

We now define for δ an element of the range of $\tilde{H}$

$$\tilde{\omega}_\delta = \inf \{\omega \epsilon \Omega : \tilde{H}(\omega) > \delta\} \qquad \ldots \ (3.5)$$

which by completeness of Ω is a r.v. Let ϵ be in the range of H_0 and H. We now pick the unique

$$\theta_\epsilon\epsilon\Omega \quad \text{such that} \quad H_0(\theta_\epsilon) = \epsilon \qquad \ldots \ (3.6)$$

and there exists a unique $\quad \omega_\epsilon\epsilon\Omega \quad$ such that $H(\omega_\epsilon) = \epsilon$. $\qquad \ldots \ (3.7)$

We now have the following theorem.

Theorem 1 : *Let assumptions* (1), (2), (3) *be true and H and $\tilde{H}_0$ be defined as in* (2.1) *and* (3.4) *respectively. Then if*

$$\Omega \supset \Phi \qquad \ldots \ (3.7.1)$$

we have $\qquad P[H(\tilde{\omega}_\delta) \geqslant \epsilon] \leqslant P[\delta \geqslant \tilde{H}_0(\theta_\epsilon)] \qquad \ldots \ (3.8)$

with the right hand side of (3.8) *being a distribution-free bound for every* $(F, G)\epsilon\mathscr{F}\times\mathscr{G}$.

Proof : It is clear by (3.1) with probability one that $\tilde{H}$ is monotone increasing on $\Omega = \Phi$ (always $\Phi \supset \Omega$). We have by the positivity of the sample space that H is strictly monotone on Ω and thus we have

$$[H(\tilde{\omega}_\delta) \geqslant \epsilon] = [\tilde{\omega}_\delta \geqslant \omega_\epsilon] \subset [\delta \geqslant \tilde{H}(\omega_\epsilon)] \qquad \ldots \ (3.8.1)$$

because $\tilde{H}(\tilde{\omega}_\delta) \leqslant \delta$. Now since $\gamma\omega_\epsilon\lambda^{-1} = \theta_\epsilon$ by (3.7.1) we have $\tilde{H}_0(\theta_\epsilon) = \tilde{H}(\omega_\epsilon)$ that $\tilde{H}_0(\theta^\epsilon)$ has a distribution independent of $\mathscr{F}\times\mathscr{G}$ is clear by assumption (3).

Remark 1 : If $\tilde{H}$ is continuous a.s. then we obtain equality between the probabilities of equation (3.8).

Remark 2 : An example of a situation when the assumption $\Omega \supset \Phi$ is satisfied occurs when Ω is a semi-group, with respect to composition, of homeomorphisms on $\mathscr{X}$ and Γ and Λ are subsets of Ω. We exhibit a trivial instance of this kind in the next section.

We now state the following theorem.

Theorem 2 : *Let assumptions* (1), (2), (3) *be true and H and H_0 be defined as in* (2.1) *and* (3.4) *respectively. Then if*

$$\text{for each} \quad (\gamma, \lambda) \, \epsilon \, \Gamma \times \Lambda \quad \text{we have} \quad \omega \geqslant \gamma \omega \lambda^{-1} \quad \text{for all } \omega \epsilon \Omega \qquad \ldots \quad (3.9)$$

then we have
$$P[H(\tilde{\omega}_\delta) \geqslant \epsilon] \leqslant P[\delta \geqslant \tilde{H}_0(\theta_\epsilon)] \qquad \ldots \quad (3.10)$$

with the right-hand side being a distribution-free bound for every $(F, G) \epsilon \mathscr{F} \times \mathscr{G}$.

Proof : It is clear by (3.1) that both H and $\tilde{H}$ are monotone on Ω. Thus we have

$$[H(\tilde{\omega}_\delta) \geqslant \epsilon] = [\tilde{\omega}_\delta \geqslant \omega_\epsilon] \subset [\delta \geqslant \tilde{H}(\omega_\epsilon)].$$

Now by (3.9) we have $\gamma \omega_\epsilon \lambda^{-1} \leqslant \omega_\epsilon$ and hence

$$\epsilon = H(\omega_\epsilon) = H_0(\gamma \omega_\epsilon \lambda^{-1}) \leqslant H_0(\omega_\epsilon)$$

but $\epsilon = H_0(\theta_\epsilon) \leqslant H_0(\omega_\epsilon)$ and since $\theta_\epsilon, \, \omega_\epsilon \, \epsilon \, \Omega$ we have $\theta_\epsilon \leqslant \omega_\epsilon$ but then a.s. $\tilde{H}(\omega_\epsilon) \geqslant \tilde{H}(\theta_\epsilon)$ and it follows that

$$[\delta \geqslant \tilde{H}(\omega_\epsilon)] \subset [\delta \geqslant \tilde{H}(\theta_\epsilon)]$$

and the theorem is proved.

Remark 3 : A useful condition which implies assumption (3.9) of Theorem 2 above, that $\lambda \omega \geqslant \gamma \omega$, is

$$\gamma \leqslant \lambda \quad \text{and either} \quad \lambda \omega \leqslant \omega \lambda \quad \text{or} \quad \gamma \omega \leqslant \omega \gamma. \qquad \ldots \quad (3.10.1)$$

In most instances this means only that one r.v. is stochastically smaller than the other and that one of the r.v.'s has a distribution which satisfies a convexity type condition.

Let us make the additional assumption

(4) $\tilde{\gamma}, \tilde{\lambda}$ are $1-1$ transformation in Γ and Λ respectively. We can in this case define

$$\hat{H}_0(\phi) = \int F\tilde{\gamma}^{-1} \phi \, dG\tilde{\lambda}^{-1}. \qquad \ldots \quad (3.11)$$

We remark that by assumption (3) being satisfied, $\hat{H}_0(\phi)$ has a distribution which depends only on $\phi \, \epsilon \, \Phi$. Let

$$\Phi_\alpha = \{\phi \epsilon \Phi : P[\hat{H}_0(\phi) \geqslant \epsilon] = \alpha\}. \qquad \ldots \quad (3.12)$$

For $\alpha \, \epsilon \, (0, 1)$, Φ_α is not empty whenever there exists $\theta_\alpha \epsilon \Omega$ such that

$$P[\hat{H}_0(\theta_\alpha) \geqslant \epsilon] = \alpha.$$

We now state the following theorem without proof :

Theorem 3 : *Let assumptions* (1), (2), (3) *and* (4), *be true with H and $\tilde{H}_0$ defined as in* (2.1) *and* (3.1) *respectively then if for each* $\phi \epsilon \Phi$ *there exists a unique largest element in* Ω, *say* ϕ_Ω, *such that* $\phi_\Omega \leqslant \phi$. $\ldots$ (3.13)

Then writing $\tilde{\omega} = (\tilde{\gamma}^{-1} \phi \tilde{\lambda})$ for $\phi \, \epsilon \, \Phi_\alpha$

$$P[H(\tilde{\omega}) \geqslant \epsilon] \leqslant P[\hat{H}_0(\phi) \geqslant \epsilon] = \alpha \qquad \ldots \ (3.14)$$

and the right hand side of (3.14) is distribution-free.

Now we define

$$\tilde{\omega}_\delta = \sup_{\phi \epsilon \Phi_\delta} (\tilde{\gamma}^{-1} \phi \tilde{\lambda}) \Omega \qquad \ldots \ (3.15)$$

as the element to be used in our estimate.

Remark 4 : In case there exists a ϕ_α such that $\tilde{\gamma}^{-1} \phi_\alpha \tilde{\lambda} = \tilde{\omega}_\alpha \epsilon \Omega$ we obtain equality between the probabilities in equation (3.14).

4. Examples

Let us first give an application of Theorem 1 and take $\mathscr{X} = (0, \infty)$, $\Omega = \{\omega : \omega(x) = \omega x, \ \omega > 0\}$, i.e., here we take the transformations ω to be scalar multiplication by a positive constant to which we give the same designation. Let the orderings $\leqslant, <$ on Ω and $\mathscr{X}$ be the same and be the usual ordering on the real line. Let F_0 and G_0 be distributions on $\mathscr{X}$ and take $\mathscr{G} = \{G_0 \omega : \omega \epsilon \Omega\}$, $\mathscr{F} = \{F_0 \omega : \omega \epsilon \Omega\}$. Without loss of generality we may assume that $EX = 1/\gamma$ and $EY = 1/\lambda$. Now we define $\tilde{F}(x) = F_0(x/\overline{X})$, $\tilde{G}(x) = G_0(x/\overline{Y})$ and these two estimates are ample for $\mathscr{F}$ and $\mathscr{G}$ respectively.

The assumptions of Theorem 1 are seen to be satisfied and θ is the unique element of Ω such that

$$\int_0^\infty F_0(\theta x) \, dG_0(x) = \epsilon \qquad \ldots \ (4.1)$$

but then $\theta_\epsilon = \gamma \omega_\epsilon \lambda^{-1}$ so a.s. we have by the corollary

$$\tilde{H}(\omega_\epsilon) = \tilde{H}_0(\theta_\epsilon) = \int_0^\infty F_0 \left(\frac{\theta_\epsilon x}{\gamma \overline{X}} \right) dG_0 \left(\frac{x}{\lambda \overline{Y}} \right) \qquad \ldots \ (4.2)$$

which must needs be tabulated at points of interest.

The integrations (4.1) and (4.2) might be difficult to perform and could require numerical integration techniques depending upon F_0 and G_0. In order to continue, let us make the mathematically convenient assumption that the densities exist and are given by

$$f_0(x) = \frac{x^{(k/2)-1} e^{-x/2}}{\Gamma(k/2) 2^{k/2}}, \ g_0(x) = e^{-x},$$

where k is a known parameter of the chi-square distribution.

Thus from (4.1) we obtain

$$\int_0^\infty F_0(\theta x) e^{-x} dx = \int_0^\infty e^{-y/\theta} f_0(y) dy = (1 + 2/\theta)^{-k/2}$$

and $\theta_\epsilon = 2/(\epsilon^{-2/k}-1)$, but fortunately this integration solves (4.2) and we see

$$\tilde{H}_0(\theta_\epsilon) = (1+2\gamma\bar{X}/\theta\lambda\bar{Y})^{-k/2}$$

from which the probability distribution can be found by a simple transformation, using tables of the F distribution, since $\gamma\bar{X}$ has a chi-square distribution with kn degrees of freedom and $2\lambda\bar{Y}$ has a chi-square distribution with $2m$ degrees of freedom where n and m are the sample sizes used in computing X and $\bar{Y}$, respectively.

5. Ample estimates of distributions

In this note we restrict ourselves to estimates $\tilde{F}$ of F which are ample. This property could be of some importance in other applications since in particular F ample for $\mathscr{F}$ tells us that the analogue of the D_n statistic, say,

$$\tilde{D} = \sup_{x\epsilon\mathscr{X}} |\tilde{F}(x)-F(x)| = \sup_{y\epsilon(0,\,1)} |\tilde{F}F^{-1}(y)-y|$$

is distribution-free with respect to $\mathscr{F}$.

Ampleness of an estimate of a distribution F is a strong property since, in many instances, it allows us to place confidence contours along the entire distribution function. Suppose $\tilde{F}$ is ample for $\mathscr{F}$, a set of distributions on the real line, and ζ is some distribution function on the unit interval which is in the range of $\tilde{F}F^{-1}$ and for some constant β we have

$$\beta = P[\tilde{F}F^{-1} \leqslant \zeta] \geqslant P[\zeta^{-1}\tilde{F} \leqslant F]$$

where $\zeta^{-1}(t) = \inf\{x : \zeta(x) \geqslant t\}$ then $\zeta^{-1}\tilde{F}$ provides a lower contour for F with β confidence.

We have ampleness within those families of distributions on the real line for which one can estimate percentiles and bounds on those estimates from observations in the region of central tendency such as normal, log-normal, exponential, etc. But we may have ampleness for distributions on spaces of higher dimension, as the following two examples show.

Let $\mathscr{G}$ be the set of continuous distributions on any partially ordered set $(\mathscr{X}, <)$ as defined in Section 2. For $y \epsilon \mathscr{X}$ define $c(.,y)$ to be the indicator (characteristic) function of the set $[Y < y]$. Then

$$\hat{G}(x) = \frac{1}{m}\sum_{j=1}^{m} c(x,Y_j) \qquad \qquad \dots \ (5.1)$$

is an ample estimate of the distribution $G \epsilon \mathscr{G}$ from the sample $(Y_1, \dots, Y_n)$ of independent observations each with distribution G. We shall call $\hat{G}$ the empiric cumulative.

Let $\mathscr{X}$ be Euclidean p-space, with $<$ the usual ordering, the elements of which we shall take to be column vectors. Let X be $\mathcal{n}(\mu, \Sigma)^*$ and let D be the non-singular matrix such that $D\Sigma D' = I$. We shall say here, for brevity, that D *diagonalizes* Σ. Then we can write $F(x) = F_0[D(x-\mu)]$ for $x \in \mathscr{X}$, where F_0 is the distribution of a $\mathcal{n}(0, 1)$ variate.

Let $\hat{\mu}$ and $\hat{\Sigma}$ be the usual maximum likelihood estimates taken from a sample of size $n > p$ and define $\hat{D}$ as the diagonalizing matrix for $\hat{\Sigma}$. It exists a.s. Now if we define

$$\sigma(x) = D(x-\mu) \qquad \qquad \text{... (5.2)}$$

and the corresponding definition of $\hat{\sigma}$ then fix $y \in \mathscr{X}$ and set

$$T = \hat{\sigma}\sigma^{-1}(y) = \hat{D}D^{-1}[y-D(\hat{\mu}-\mu)], \qquad \text{... (5.3)}$$

that this has a distribution which is the same for all μ, Σ is sufficient for ampleness. But clearly $D(\hat{\mu}-\mu)$ is $\mathcal{n}\left(0, \dfrac{1}{n} I\right)$ which is sufficient that the vector in square brackets in (5.3) satisfies the condition.

Set $\hat{B} = \hat{D}\hat{D}^{-1}$. We show that this has a distribution independent of μ, Σ also. From (5.2) we have

$$\hat{B}D \hat{\Sigma} D'\hat{B}' = 1. \qquad \qquad \text{... (5.4)}$$

But $nD \hat{\Sigma} D'$ is distributed as $\sum\limits_{i=1}^{n-1} Z_i Z_i'$ where Z_i is $\mathcal{n}(0, 1)$ independent of Z_j $(i \neq j)$.

Hence, except for zero probability; $\hat{B}$ diagonalizes a random matrix which has a distribution independent of μ, Σ. Note that $T'T$ is, except for a scale factor, Hotelling's T^2 with non-centrality factor $y'y/n$.

6. Some remarks on properties of the estimate H

Let us set $\bar{F} = E\tilde{F}$ with similar meaning for $\bar{G}$ and $\bar{H}$. By applying the Fubini theorem and integrating by parts we have for ω

$$\bar{H}(\omega) = \int \bar{F}\omega d\bar{G}$$

and
$$H(\omega) - \bar{H}(\omega) = \int F\omega d(G-\bar{G}) + \int (F\omega - \bar{F}\omega)dG.$$

Remark 5 : $\tilde{H}$ is an unbiased estimate of H on Ω if $\tilde{F}$ and $\tilde{G}$ are unbiased estimates of F and G, respectively.

The converse is not true without additional assumptions. For the real line, unbiasedness of $\tilde{H}$ on the entire set of $1-1$ order preserving maps of $\mathscr{X}$ onto itself is sufficient for unbiasedness of $\tilde{F}$ and $\tilde{G}$ when one imposes the conditions that $\bar{F}$ and $\bar{G}$ are continuous.

---

* i.e., X is a normal vector with mean vector μ and variance-covariance matrix Σ.

We say that $\tilde{F}_n$ is consistent for F whenever

$$\sup_{\mathscr{X}} \ |\tilde{F}_n - F| \ P \xrightarrow{P} 0,$$

where, as before, the subscript n refers to the sample size used to obtain the estimate. We shall refer to consistency exclusively in this sense.

Now

$$\tilde{H}(\omega) - H(\omega) = \int (G - \tilde{G}) d\tilde{F}\omega + \int (\tilde{F}\omega - F\omega) dG.$$

Thus it follows that $\ |\tilde{H}(\omega) - H(\omega)| \leqslant \sup_{\mathscr{X}} \ |\tilde{F} - F| + \sup_{\mathscr{X}} \ |\tilde{G} - G|.$

Remark 6 : If $\tilde{F}_n$ and $\tilde{G}_m$ are consistent for F and G, respectively, then $\tilde{H}_{mn}$ is consistent for H uniformly on Ω as $1/n + 1/m - 0$.

Remark 7 : If $\hat{F}_n$ and $\hat{G}_m$ are the empiric cumulatives, then H_{mn}, for each $\omega \in \Omega$, is consistent, unbiased and asymptotically normal with $1/n + 1/m \to 0$.

This follows from the known behaviour of U-Statistics.

If we assume that $\tilde{F}$ is consistent and $\hat{G}$ is the empiric cumulative, we should obtain asymptotic normality of $\tilde{F}$ "stabilizes" rapidly enough. We do not attempt to find the weakest of such conditions, but we have the following remark.

Remark 8 : If $\hat{G}$ is the empiric cumulative, then the function H defined by

$$H_{mn}(\omega) = \frac{1}{m} \sum_{j=1}^{m} \tilde{F}_n(\omega Y_j)$$

is a consistent estimator of H and asymptotically normal if $1/n + 1/m \to 0$ in such a way that $\sqrt{m} \sup |\tilde{F}_n - F| \xrightarrow{P} 0.$

Proof : Let $Z_m = \frac{1}{\sqrt{m}} \sum_{j=1}^{m} [F(\omega Y_j) - H(\omega)]$ and define $\zeta_{m,n}$ by the equation $\sqrt{m}[\tilde{H}_{mn}(\omega) - H(\omega)] = Z_n + \zeta_{m,n}.$ That Z_m is asymptotically normal is clear, and the result follows if $\zeta_{m,n} \xrightarrow{P} 0$, which is guaranteed by the hypothesis.

7. AN APPLICATION TO RELIABILITY THEORY

Let $\mathscr{X} = (0, \infty)$ and let $Y(\omega)$ be the demand time for a particular equipment in a specified environment during an inspection period of length ω. Let X be the life length of this equipment under continued use in that environment.

We want the maximum ω such that for given $\epsilon < 0$

$$P[X < Y(\omega)] \leqslant \epsilon.$$

We particularize by making the assumption that $Y(\omega)$ denotes multiplication and without loss of generality we suppose that a scaling factor has been introduced so that $\omega \epsilon (0, 1)$. Now

$$H(\omega) = P[X < \omega Y] = \int F(\omega y) dG(y)$$

where we assume that X and Y are independent r.v.'s with continuous distributions F and G respectively.

Now if F is a distribution with derivative f then the function γ' defined by

$$\gamma'(x) = \frac{f(x)}{1-F(x)} \quad \text{for } x > 0 \qquad \ldots \quad (7.1)$$

is called the *failure rate* (or hazard rate).

A common and intuitively appealing assumption concerning life length distributions is

(A) the failure rate of X exists and is non-decreasing.

We also assume

(B) X is stochastically larger than Y, i.e., $F \leqslant G$.

But, further, without loss of generality we assume

$$\frac{1}{2} > \varepsilon > 0 \text{ is given and } P[X < Y] > \varepsilon. \qquad \ldots \quad (7.2)$$

Let

$$\gamma(x) = \int_0^x \gamma'(t)dt \quad \text{for } x > 0:$$

Then if we set

$$F_0(x) = G_0(x) = 1-e^{-x}, \, x > 0,$$

from (7.1) we see that $G(x) = G_0\lambda(x)$.

It is clear that (B) implies $\gamma \leqslant \lambda$ and we now show that (A) implies $\gamma\omega \leqslant \omega\gamma$ and thus (3.10.1) implies assumption (3.9) of Theorem 2. To see this, note that γ must be convex and hence $\gamma(\omega x) \leqslant \omega\gamma(x)$ for $\omega \in (0, 1)$ since $\gamma(0) = 0$. This is precisely our condition.

We now seek the unique θ such that

$$\int_0^\infty F_0(\theta x) \, dG_0(x) = \varepsilon. \qquad \ldots \quad (7.3)$$

This may be easily integrated directly or by comparison with (4.1) seen to be $\theta_\varepsilon = \dfrac{\varepsilon}{1-\varepsilon}$.

We now choose $\hat{F}, \hat{G}$ to be the empiric cumulatives as defined in (5.1) (which are ample). Hence we need tabulations of the statistic

$$\hat{H}_0(\theta) = \int \hat{F}F^{-1}\theta d\hat{G}G^{-1} = \frac{1}{mn} \sum_{i=1}^{n} \sum_{j=1}^{m} c(\theta U_j, V_i)$$

where $F(X_i) = V_i$, $i = 1, ..., n$, and $G(Y_j) = U_j$ for $j = 1, ..., m$ are all independent r.v.'s uniform on $(0, 1)$.

Consider the events

$$A_k = [V_{(k)} < \theta < V_{(k+1)}], \quad k = 0, ..., n,$$

where $V_{(1)}, ..., V_{(n)}$ are the ordered observations of V with $V_{(0)} = 0$, $V_{(n+1)} = 1$. These events A_k form a disjunct partition of the sample space.

It is clear that $\hat{H}_0(1) = U_{m,n}$ is the well-known Mann-Whitney statistic based on samples of size m and n. Now let $S_{mn}(\theta) = mnH_0(\theta)$, and it follows that

$$S_{mn}(\theta)|A_k = U_{mk} \quad \text{for } k = 0, 1, ..., n.$$

With the convention that $U_{m0} = 0$ with probability one, we obtain

$$P[S_{mn}(\theta) \leqslant t] = \sum_{k=0}^{n} P[S_{mn}(\theta) \leqslant t \,|\, A_k]P[A_k]$$

$$= \sum_{k=0}^{n} b(k; n, \theta)P[U_{mk} \leqslant t], \quad t = 0, 1, \ldots, mn.$$

We know that $ES_{mn}(\theta) = \dfrac{mn}{2}\,\theta$ and since we have

$$EU_{mn}^2 = \frac{nm}{12}\,(n+m+1)+(nm/2)^2,$$

it follows by (7.4) directly that

$$ES_{mn}^2(\theta) = \sum_{k=0}^{n} b(k : n, \theta)EU_{mk}^2$$

$$= \frac{m(m+1)n\theta}{12} + \left(\frac{m}{12} + \frac{m^2}{4}\right)[n(n-1)\theta^2 + n\theta].$$

Hence we obtain the variance of $S_{mn}(\theta)$ as

$$\mathrm{var}(S_{mn}(\theta)) = \frac{mn\theta}{6}\left[2m - \frac{3m\theta}{2} + \frac{(n-1)\theta}{2} + 1\right].$$

The results of the preceding section show that $S_{mn}(\theta)$ has asymptotically a normal distribution with the above mean and variance; however, for small sample sizes and small θ the normal approximation may not be of sufficient accuracy. In the following paragraph we give a few formulae in the range of $t = mn\delta$ small. This is near the region of interest since α is small.

Let $P_{mn}(t) = P[U_{mn} \leqslant t]$. Now, from results and recurrence formulae given by Mann and Whitney (1947), one can obtain the following :

For $\qquad m \geqslant 1, k \geqslant 0,$

$$P_{mk}(0) = 1 \Big/ \binom{m+k}{k} \qquad\qquad \text{for all } k \geqslant 0,$$

$$P_{mk}(1) = \begin{cases} 1 & k = 0, \\ 2 \Big/ \binom{m+k}{k} & k \geqslant 1, \end{cases}$$

$$P_{mk}(2) = \begin{cases} 1 & \text{if } m \leqslant 2 \text{ or } k = 0, 1, \\ 4 \Big/ \binom{m+k}{k} & \text{if } m \geqslant 2 \text{ and } k \geqslant 2, \end{cases}$$

$$P_{mk}(3) = \begin{cases} 1 & \text{if } m \leqslant 3 \text{ or } k = 0, 1, \\ (4+k) \Big/ \binom{m+k}{k} & \text{if } m \geqslant 3 \text{ and } k \geqslant 2, \end{cases}$$

and for greater values of t the numerators are functions of higher powers of k.

We have the expression

$$F[S_{mn}(\theta) \leqslant 0] = \sum_{k=0}^{n} \frac{b(k : n, \theta)}{\binom{m+k}{k}} = \frac{E(m+n, m, \theta)}{\binom{m+n}{n}\theta^m}, \qquad \ldots \quad (7.5)$$

where the notation E is that used in the Harvard Tables of the Binomial Distribution.* Through the use of these tables, for moderate values of m, n and θ, the probability (7.5) can be calculated and for very small values of θ the first terms of the expansion given can be used for adequate approximation.

Similarly :

$$P[S_{mn}(\theta) \leqslant 1] = (1-\theta)^n + \frac{2E(m+n, m+1, \theta)}{\binom{m+n}{m}\theta^m},$$

$$\text{for } n \geqslant 1, m \geqslant 1,$$

and

$$P[S_{mn}(\theta) \leqslant 2] = (1-\theta)^n + n\theta(1-\theta)^{n-1} + \frac{E(m+n, m+2, \theta)}{\binom{m+n}{m}\theta^m},$$

$$\text{for } n \geqslant 1, m \geqslant 2.$$

We now return to the reliability problem. From the formulae preceding (or others similarly developed) we discover, through tabulation, the α-th percentile of the r.v. $S_{mn}(\theta)$, thus we find δ_a, dividing by mn, such that

$$P[\hat{H}_0(\theta) \leqslant \delta_a] = \alpha.$$

Then we set $\qquad \hat{\omega}_\delta = \inf \{\omega \epsilon(0, 1) : \hat{H}(\omega) \geqslant \delta_a\}$

which is the empiric derating percentage sought. We know by the assumptions (A) and (B) that the theorem applies and thus

$$P[H(\hat{\omega}_\delta) \geqslant \epsilon] \leqslant \alpha.$$

In this case, for example, if we chose $\delta = 0$ we would find that

$$\hat{\omega}_0 = \frac{\min X_i}{\max Y_j}$$

which is intuitively appealing.

REFERENCES

BIRNBAUM, Z. W. (1956) : On a use of the Mann-Whitney statistic. *Proc. Third Berkeley Symposium*, 13-17.

BIRNBAUM Z. W. and McCARTY, R. C. (1958) : A distribution-free upper confidence bound for $P[Y < X]$, based on independent samples of X and Y. *Ann. Math. Stat.*, 29, 558-562.

MANN H. B. and WHITNEY, D. R. (1947) : On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Stat.*, 18, 50-60.

Paper received : May, 1961.

* *Tables of the Cumulative Binomial Probability Distribution*, Harvard University Press, 1955.

STUDENTISATION OF TWO-STAGE SAMPLE MEANS FROM NORMAL POPULATIONS WITH UNKNOWN VARIANCES[1]

II. CONFIDENCE ESTIMATION FOR THE MEAN OF A STRATIFIED POPULATION

By HAROLD RUBEN[2]

Columbia University

SUMMARY. The two-stage sampling procedure discussed in the first paper[3] of the present series is used to obtain confidence intervals of predetermined length and confidence coefficient for the mean of a stratified population with normal components. The efficiency of the procedure in relation to the confidence estimation with fixed sample sizes, when the variances of the component strata are known, is discussed briefly.

For samples of fixed size the estimation of the overall mean μ of a stratified (weighted) population with r components having means and variances μ_i and σ_i^2 and weights $p_i(i = 1, 2, ..., r)$, the reliability of the unbiassed minimum variance estimator involves the (generally unknown) σ_i^2. Accordingly, it has frequently been suggested (e.g. Cochran, 1953) that pilot samples be first drawn to gain information about the σ_i^2 (and incidentally also to enable the optimum allocation of sampling among the strata to be determined). This suggestion tallies well with the two-stage sampling procedure discussed in this paper.

We shall consider here briefly the problem of obtaining confidence intervals of *predetermined* length $2a$ and predetermined confidence coefficient $1-\alpha$ for

$$\mu = \sum_1^r p_i \mu_i \qquad \qquad \cdots \ (1)$$

when the components of the population are normal. The appropriate random variable is clearly $\sum_1^r p_i(\bar{x}_i - \mu_i)$, and the sampling procedure $S(n_{01}, ..., n_{0r}; k_1, ..., k_r)$ is used, where the procedure S may be described as follows (Ruben, 1962). Preliminary independent random samples of size $n_{0i}(i = 1, 2, ..., r)$ are drawn from the r populations, the sample observations being x_{ij} $(i = 1, 2, ..., r; j = 1, 2, ..., n_{0i})$. Further, independent random samples of size $n_i - n_{0i}$ are drawn from the r populations according to the rule

$$n_i = \max(\{k_i^2 s_{0i}^2\}, n_{0i}) \quad (i = 1, 2, ..., r), \qquad \cdots \ (2)$$

where

$$s_{0i}^2 = (n_{0i}-1)^{-1} \left[\sum_{j=1}^{n_{0i}} x_{ij}^2 - \left(\sum_{j=1}^{n_{0i}} x_{ij} \right)^2 \Big/ n_{0i} \right], \qquad \ldots \ (2.1)$$

and $\{c\}$ denotes the smallest integer not exceeded by c. The x_i are the sample means based on all the observations,

$$\bar{x}_i = \sum_{j=1}^{n_i} x_{ij}/n_i. \qquad \ldots \ (2.2)$$

We have (Ruben, 1962, equ. (3.11)) that in the limit as $\sigma_i \to \infty$,

$$z = \sum_1^r p_i(\bar{x}_i - \mu_i) \sim \sum_1^r p_i \, t_{\nu_{0i}}/k_i, \qquad \ldots \ (3)$$

where the $t_{\nu_{0i}}$ are independent Student variables with ν_{0i} degrees of freedom, and

$$\nu_{0i} = n_{0i}-1.$$

The distribution of z is not available, and for simplicity we choose

$$k_i/p_i = \lambda \quad (i = 1, 2, \ldots, r), \qquad \ldots \ (4)$$

where λ has to be chosen appropriately. Further, we choose, again for simplicity, $n_{01} = n_{02} = \cdots = n_{0r} = n_0$ (say).

On setting

$$z = \frac{1}{\lambda} \left(\frac{r\nu_0}{\nu_0-2} \right)^{\frac{1}{2}} w, \qquad \ldots \ (5)$$

where $\nu_0 = n_0-1$ and

$$w = \left(\frac{\nu_0-2}{r\nu_0} \right)^{\frac{1}{2}} \sum_1^r t_{\nu_{0i}} \qquad \ldots \ (6)$$

(w is a standardised variate with zero mean and unit variance), we find, on using the Cornish-Fisher expansion (1937), the following approximate relation between the percentile points of z and ξ, where ξ is normal with zero mean and unit variance,

$$z_{\alpha/2} = \frac{1}{\lambda} \left(\frac{r\nu_0}{\nu_0-2} \right)^{\frac{1}{2}} \cdot [\xi_{\alpha/2} + (\xi_{\alpha/2}^3 - 3\xi_{\alpha/2})/(4r\nu_0)$$

$$+ \{5\xi_{\alpha/2}^5 + (96r-104)\xi_{\alpha/2}^3 + (219-288r)\xi_{\alpha/2}\}/(96r^2\nu_0^2)], \qquad \ldots \ (7)$$

terms of order $1/\nu_0^3$ being disregarded. Since in order to meet the specifications we require $z_{\alpha/2} = a$, this, taken in conjunction with (4), yields

$$k_i = \left(\frac{r\nu_0}{\nu_0-2} \right)^{\frac{1}{2}} \frac{p_i}{a} [\xi_{\alpha/2} + (\xi_{\alpha/2}^3 - 3\xi_{\alpha/2})/(4r\nu_0)$$

$$+ \{5\xi_{\alpha/2}^5 + (96r-104)\xi_{\alpha/2}^3 + (219-288r)\xi_{\alpha/2}\}/(96r^2\nu_0^2)], \qquad \ldots \ (8)$$

to the same order of accuracy as $z_{\alpha/2}$ in (7).*

---

* Detailed examination of higher order terms in the series expansion for k_i indicates that the disregarding of 3rd order terms is likely to affect only the 4th decimal place, even for moderate ν_0.

The value of Q_m, the conditional probability for fixed sample sizes m_i, corresponding to the region R in the $(\bar{x}_1, \ldots, \bar{x}_r)$-plane defined by

$$R : \left| \sum_1^r p_i(\bar{x}_i - \mu_i) \right| < a \qquad \ldots \ (9)$$

is

$$2\Phi[a(\sum_1^r p_i^2\sigma_i^2/m_i)^{-\frac{1}{2}}] - 1, \qquad \ldots \ (10)$$

where $\Phi(\cdot)$, as usual, denotes the distribution function of a normal random variable with zero mean and unit variance. Since Q_m is decreasing in each of the σ_i, it follows from the results of the first paper (Ruben, 1962) of this series that the present estimation procedure is conservative in the sense that the true confidence coefficient exceeds $1 - \alpha$.

Finally, we investigate the efficiency of the procedure by comparing its average sample size with the *fixed* sample size required in order to meet the given requirements when the σ_i^2 are known. Under these circumstances (using the standardised normal variate for the estimation of μ),

$$\frac{a}{\left(\sum_1^r p_i^2\sigma_i^2/n_i \right)^{\frac{1}{2}}} = \xi_{\alpha/2}, \qquad \ldots \ (11)$$

n_i being the sample size for the i-th stratum. The minimisation of

$$n = \sum_1^r n_i \qquad \ldots \ (12)$$

subject to (11) leads to

$$n_i^* = \xi_{\alpha/2}^2 \left(\sum_1^r p_j\sigma_j \right) p_i\sigma_i / a^2, \qquad \ldots \ (13)$$

whence the minimal n is

$$n^* = \xi_{\alpha/2}^2 \left(\sum_1^r p_i\sigma_i \right)^2 / a^2. \qquad \ldots \ (14)$$

On the other hand, for the procedure $S(n_0, \ldots, n_r; k_1, \ldots, k_r)$, provided the σ_i^2 are not excessively small,

$$En = \sum_1^r En_i \sim \sum_1^r k_i^2\sigma_i^2 = w_{\alpha/2}^2 \frac{r\nu_0}{\nu_0 - 2} \left(\sum_1^r p_i^2\sigma_i^2 \right) / a^2. \qquad \ldots \ (15)$$

The efficiency is therefore

$$\frac{n^*}{En} = \frac{\nu_0 - 2}{\nu_0} \left(\frac{\xi_{\alpha/2}}{w_{\alpha/2}} \right)^2 \times \frac{1}{r} \frac{\left(\sum_1^r p_i\sigma_i \right)^2}{\sum_1^r p_i^2\sigma_i^2}. \qquad \ldots \ (16)$$

Since the second term on the right of (16) < 1, unless either the $p_i\sigma_i$ are equal or $r = 1$, when it equals 1,

$$\text{Efficiency} \leqslant \frac{\nu_0 - 2}{\nu_0} \left(\frac{\xi_{\alpha/2}}{w_{\alpha/2}} \right)^2. \qquad \ldots \ (17)$$

(17) gives an upper bound to the efficiency. Supposing, however, we do not allow ourselves the degree of freedom available in the fixed sampling size procedure (when the σ_i^2 are known) but rather consider the class of fixed sampling procedures characterised by the property

$$\frac{n_1}{En_1} = \cdots = \frac{n_r}{En_r} = \frac{\lambda'^2}{\lambda^2}, \qquad \cdots \; (18)$$

i.e., we consider only the case where

$$n_i = \lambda'^2 p_i^2 \sigma_i^2 \quad (i = 1, 2, ..., r), \qquad \cdots \; (19)$$

where λ' is yet to be determined. From (11),

$$\lambda'^2 = r(\xi_{\alpha/2}/a)^2$$

and the total fixed sample size now becomes

$$n = \sum_1^r n_i = r(\xi_{\alpha/2}/a)^2 \sum_1^r p_i^2 \sigma_i^2. \qquad \cdots \; (20)$$

The 'conditional' efficiency of the two-stage sampling estimation procedure in relation to the class of fixed sampling size procedures, wherein the fixed sample sizes of the populations are allocated in the same ratio as the average sample sizes in the two-stage procedure, is then

$$\overline{\text{Eff}} = \frac{n}{\sum_1^r En_i} = \frac{v_0-2}{v_0} \left(\frac{\xi_{\alpha/2}}{w_{\alpha/2}}\right)^2$$

$$= \frac{v_0-2}{v_0} \left[1 + \frac{\xi_{\alpha/2}^2-3}{4rv_0} + \frac{5\xi_{\alpha/2}^4+(96r-104)\xi_{\alpha/2}^2+(219-288r)}{96r^2v_0^2} \right]^{-2}, \qquad \cdots \; (21)$$

independently of the σ_i^2 (cf. Ruben, 1962). This is also the upper bound to the efficiency, in the original and wider sense, as given in (16). The conditional efficiency approaches 1 rapidly as $v_0 \to \infty$.

REFERENCES

COCHRAN, W. G. (1953) : *Sampling Techniques*, John Wiley & Sons, New York.

CORNISH, E. A. and FISHER, R. A. (1937) : Moments and cumulants in the specification of distributions. *Rev. Inst. Int. Statist.*, **4**, 11.

RUBEN, H. (1962) : Studentisation of two-stage sample means from normal populations with unknown variances. I. General theory and application to the confidence estimation and testing of the difference in population means. *Sankhyā*, Series A, **24**, Part 2, 157-180.

Paper received : December, 1960.

STUDENTISATION OF TWO-STAGE SAMPLE MEANS FROM NORMAL POPULATIONS WITH UNKNOWN VARIANCES[1]

III. JOINT CONFIDENCE ESTIMATION OF A SET OF MEANS

By HAROLD RUBEN[2]

Columbia University

SUMMARY. Confidence regions of predetermined dimensions and confidence coefficient are obtained for the joint estimation of the means of a finite set of unconnected and unknown normal populations. The confidence estimation procedure is shown to be conservative in character, and its efficiency in relation to the corresponding estimation procedure with fixed sample sizes, when the variances of the populations are known, is discussed briefly.

The paper concludes with a short discussion of the possibility of increasing the efficiency of the two-stage sampling procedure (utilised in the current paper as well as in the preceding two papers[3] of this series), and of some other matters.

DISCUSSION OF THE PROCEDURE

Given r separate normal populations with unknown means and variances μ_i, σ_i^2, respectively, we wish to obtain confidence regions for $\mu = (\mu_1, \mu_2, ..., \mu_r)$ of *fixed* dimensions and confidence coefficient $1-\alpha$. We propose to use

$$y = \prod_1^r \left\{ \left(1 + \frac{k_i^2}{\nu_{0i}} (\bar{x}_i - \mu_i)^2 \right)^{-\frac{1}{2}(\nu_{0i}+1)} \right\} \qquad \ldots \text{(1)}$$

as the basic random variable and $S(n_{01}, ..., n_{0r}; k_1, ..., k_r)$ [Ruben, 1962] as the sampling procedure, with $\nu_{0i} = n_{0i}-1$. Observe that the random variable chosen is proportional to the joint density function of the $\bar{x}_i$.[4]

The procedure S may be described as follows: Preliminary independent random samples of size n_{0i} $(i = 1, 2, ..., r)$ are drawn from the r populations, the sample observations being x_{ij} $(i = 1, 2, ..., r; \ j = 1, 2, ..., n_{0i})$. Further independent random samples of size $n_i - n_{0i}$ are drawn from the r populations according to the rule

$$n_i = \max \left(\{k_i^2 s_{0i}^2\}, n_{0i} \right) \qquad (i = 1, 2, ..., r), \qquad \ldots \text{(2)}$$

where

$$s_{0i}^2 = (n_{0i}-1)^{-1} \left[\sum_{j=1}^{n_{0i}} x_{ij}^2 - \left(\sum_{j=1}^{n_{0i}} x_{ij} \right)^2 \middle/ n_{0i} \right] \qquad \ldots \text{(2.1)}$$

and $\{c\}$ denotes the smallest integer not exceeded by c. The $\bar{x}_i$ are the sample means based on all the observations,

$$\bar{x}_i = \sum_{j=1}^{n_i} x_{ij}/n_i . \qquad \ldots \text{(2.2)}$$

---

[4] One would perhaps intuitively expect that this variate will generate confidence regions of minimum volume for given confidence coefficient and total average sample size. Some independent support is lent to the choice of y by the fact that the use of the Neyman-Pearson likelihood ratio for testing the hypothesis $\mu_i = \mu_i^0$ $(i = 1, 2, ..., r)$ with fixed sample sizes leads to the same quantity with μ_i replaced by μ_i^0.

From a result of a previous paper (Ruben, 1962) the limiting distribution of $y(\sigma_i \to \infty)$ is the distribution of

$$\prod_{1}^{r} \left\{ \left(1+\frac{t_{\nu_{0i}}^{2}}{\nu_{0i}} \right)^{-\frac{1}{2}(\nu_{0i}+1)} \right\}, \qquad \ldots \ (3)$$

where the $t_{\nu_{0i}}$ are independent Student variables with ν_{0i} degrees of freedom, and since $(1+t_{\nu_{0i}}^{2}/\nu_{0i})^{-1}$, say y_i, is a Beta variate with density function

$$\frac{\Gamma\left(\frac{\nu_{0i}+1}{2}\right)}{\Gamma\left(\frac{\nu_{0i}}{2}\right)\sqrt{\pi}} \ y_i^{\frac{1}{2}\nu_{0i}-1}(1-y_i)^{-\frac{1}{2}} \qquad (0 < y_i < 1), \qquad \ldots \ (4)$$

y is distributed as the weighted product of r such independent Beta variables. The distribution of the product of Beta variables is unknown and an approximation for the distribution of y is therefore called for. An excellent approximation is provided by noting that the distribution of

$$-2lny = -\sum_{i=1}^{r} (\nu_{0i}+1)lny_i \qquad \ldots \ (5)$$

is asymptotically ($\nu_{0i} \to \infty$) that of a χ^2 with r degrees of freedom. This suggests that, for finite ν_{0i}, $-2lny$ may be regarded approximately as a Type III Pearson variable, and this may be verified by inspection of the cumulants of $-2lny$. (For analogous approximations see, for instance, Bartlett, 1954.) The latter cumulants are obtained by noting that the cumulant generating function of lny_i is

$$ln\Gamma\left(\frac{\nu_{0i}+1}{2}\right) - ln\Gamma\left(\frac{\nu_{0i}}{2}\right) + ln\Gamma\left(\frac{\nu_{0i}}{2}+jt\right) - ln\Gamma\left(\frac{\nu_{0i}+1}{2}+jt\right) \qquad \ldots \ (6)$$

$(j=+\sqrt{-1})$, whence the cumulants, $\kappa_s(-2lny)$, may be obtained in terms of the Pentagamma functions $\psi^{(s-1)}(.)$ (British Association, 1931), as follows :

$$\kappa_s(-2lny) = (-)^s \sum_{i=1}^{r} (\nu_{0i}+1)^s \left\{ \psi^{(s-1)}\left(\frac{\nu_{0i}}{2}\right) - \psi^{(s-1)}\left(\frac{\nu_{0i}+1}{2}\right) \right\}, \qquad \ldots \ (7)$$

where $\quad \psi^{(0)}(x) \equiv \psi(x) = \frac{d}{dx} ln\Gamma(x), \ \psi^{(s-1)}(x) = \frac{d^{s-1}}{dx^{s-1}} \psi(x) \ (s = 2, 3, \ldots). \quad \ldots \ (7.1)$

Asymptotic developments of the Pentagamma functions in terms of negative powers of the ν_{0i} yields corresponding asymptotic developments for the cumulants of $-2lny$ in equ. (7), and subsequent inspection of these cumulants in series form confirms the closeness of approximation of $-2lny$ to a Type III Pearson variable. We give here only the first two cumulants, disregarding terms of higher order than the third:

$$\kappa_1(-2lny) = r+ \frac{3}{2} \sum_{1}^{r} 1/\nu_{0i} + \frac{1}{2} \sum_{1}^{r} 1/\nu_{0i}^{2} + \frac{1}{4} \sum_{1}^{r} 1/\nu_{0i}^{3}, \qquad \ldots \ (8)$$

$$\kappa_2(-2lny) = 2r+6 \sum_{1}^{r} 1/\nu_{0i} + 6 \sum_{1}^{r} 1/\nu_{0i}^{2}. \qquad \ldots \ (9)$$

The sought for approximation is then that

$$\frac{-2lny}{1+(3/2r)\sum_{1}^{r} 1/\nu_{0i}} \qquad \ldots \ (10)$$

is distributed as a chi-square on r degrees of freedom, and with this approximation the confidence regions are provided by the relation

$$\prod_1^r \left\{ (1+(k_i^2/\nu_{0i})(\mu_i-x_i)^2)^{-\frac{1}{2}(\nu_{0i}+1)} \right\} > \exp\left[-\tfrac{1}{2}(1+(3/2r)\sum_1^r 1/\nu_{0i})\chi_{r;\,\alpha}^2 \right], \qquad \ldots \quad (11)$$

where $\chi_{r;\,\alpha}^2$ is the $100\alpha\%$ significance point of a chi-square with r degrees of freedom. The confidence regions are ovaloids with semi-axes

$$l_i = (\nu_{0i}^{1/2}/k_i)\left[\exp\left\{ \frac{(1+(3/2r)\sum_1^r 1/\nu_{0i})\,\chi_{r;\,\alpha}^2}{1+\nu_{0i}} \right\} - 1 \right]^{\frac{1}{2}} \quad (i=1,2,\ldots,r), \qquad \ldots \quad (12)$$

and are hardly distinguishable from ellipsoidal regions* with lengths of semi-axes

$$l_i' = (1/k_i)\,[(1+(3/2r)\sum_1^r 1/\nu_{0i})\,\chi_{r;\,\alpha}^2]^{\frac{1}{2}}. \qquad \ldots \quad (13)$$

Equation (12) enables the lengths of the axes to be controlled by an appropriate choice of the k_i.

The value of the conditional probability for *fixed* sample sizes m_i corresponding to (11) is

$$\Pr\left\{ \prod_1^r \left(1+ \frac{k_i^2}{\nu_{0i}}\,\frac{\sigma_i^2\xi_i^2}{m_i} \right)^{-\frac{1}{2}(\nu_{0i}+1)} > \exp\left[-\tfrac{1}{2}\left(1+(3/2r)\sum_1^r 1/\nu_{0i}\right)\chi_{r;\,\alpha}^2 \right] \right\}, \cdots \quad (14)$$

where the ξ_i are independent standardised normal variables, and is decreasing in the σ_i. (The effect of successively increasing the σ_i is to produce a decreasing sequence of ovaloids.) It then follows from the results of a previous paper (Ruben, 1962) that the present estimation procedure is conservative in character, and the true confidence coefficient exceeds $1-\alpha$.

To investigate the *efficiency* of the procedure, note that when the σ_i are *known* confidence regions for μ with confidence coefficient $1-\alpha$ are the random ellipsoidal regions in the space of $(\mu_1,\ldots,\mu_r)$,

$$\sum_1^r n_i(\mu_i-x_i)^2/\sigma_i^2 < \chi_{r;\,\alpha}^2 \qquad \ldots \quad (15)$$

where the sample sizes n_i are fixed. To obtain l_i' $(i=1,2,\ldots,r)$ as the lengths of the semi-axes, the n_i must be chosen so that

$$n_i = (\chi_{r;\,\alpha}^2\sigma_i^2)/l_i'^2, \qquad \ldots \quad (16)$$

the total number of items sampled being then

$$n = \sum_1^r n_i = \chi_{r;\,\alpha}^2 \sum_1^r \sigma_i^2/l_i'^2. \qquad \ldots \quad (17)$$

On the other hand, when the σ_i are unknown the *average* total sample size necessary to meet the same requirements is

$$En = \sum_1^r En_i \sim \sum_1^r k_i^2\,\sigma_i^2 = \left(1+(3/2r)\sum_1^r 1/\nu_{0i}\right)\chi_{r;\,\alpha}^2 \sum_1^r \sigma_i^2/l_i'^2. \qquad \ldots \quad (18)$$

The efficiency of the two-stage procedure is then

$$\frac{n}{En} = (1+(3/2r)\sum_1^r 1/\nu_{0i})^{-1}. \qquad \ldots \quad (19)$$

---

* Ellipsoidal confidence regions would be obtained if the variate $\sum_1^r k_i^2\,(x_i-\mu_i)^2$ were used as the starting point rather than y. Similarly, confidence regions of various shapes can be obtained by taking appropriate functions of the $x_i-\mu_i$.

We note that the efficiency is (approximately) a function only of the harmonic average of the preliminary sample degrees of freedom.

Concluding remarks

We conclude the present series of papers on the sampling scheme $S(n_{01}, ..., n_{0r}; k_1, ..., k_r)$ with some further remarks. These are essentially of the same character as those of a previous paper (Ruben, 1961, Section 6) and will therefore be referred to quite briefly. The remarks of the earlier paper related to the case of equal variabilities for which σ is scalar, while here, where the variabilities are totally unknown, σ is a vector, but the validity of the remarks is not affected by this difference and the discussion carries over with suitable and obvious modifications.

(i) It follows from equ. (3.12) of the first paper of the present series (Ruben, 1962), that the test procedures discussed in the series remain valid and are furthermore conservative even if σ has some prior distribution, provided $P(R|\mu, \sigma)$, the probability that the vector of the sample means falls in the critical region R of the test for given μ and σ, is monotone in each of the σ_i. Similarly, the confidence estimation procedures remain valid and are conservative even if μ and σ have a joint prior distribution, provided $P(R|\sigma)$[1] is decreasing in each of the σ_i. (R is here a region in the space of the $\bar{x}_i$ which is used to derive a confidence set for μ).

(ii) For purposes of inference all probabilities have been evaluated in this and previous articles as limiting probabilities, i.e. we have assumed infinitely high σ_i for the sake of safety. In practice, however, safe upper bounds would generally be available for the σ_i, and use of these upper bounds in $P(R|\mu, \sigma)$ will improve on the confidence probabilities or risks of errors, etc.

(iii) Stein's two-stage sampling procedure $S(n_0, k)$ for the mean of a single normal population with unknown variance has been criticised (for references, see Ruben, 1961) on the grounds that information relevant to an assessment of the variance is thrown away by the failure to utilise the observations in the second-stage sample for this purpose. The criticism obviously carries over to the present procedure $S(n_{01}, ..., n_{0r}; k_1, ..., k_r)$ for the means of several populations. Nevertheless, it may to some extent be met by incorporating estimates of the σ_i, obtained from the totality of observations, in the probabilities $P(R|\mu, \sigma)$.[2] This will have the effect of increasing the efficiency of the procedure.

References

Bartlett, M. S. (1954) : A note on the multiplying factors for various χ^2 approximations. *J. Roy. Stat. Soc.*, B, **16**, 296–298.

British Association (1931) : *Mathematical Tables*, Vol. 1, Cambridge University Press.

Ruben, H. (1961) : Studentisation of two-stage sample means from normal populations with unknown common variance. *Sankhyā*, Series A, **23**, 231—250.

———— (1962) : Studentisation of two-stage sample means from normal populations with unknown variances. I. General theory and application to the confidence estimation and testing of the difference in population means. *Sankhyā*, Series A, **24**, 157–180.

[1] In the estimation situation, $P(R|\mu, \sigma)$ cannot involve μ and may therefore be represented by $P(R|\sigma)$.

[2] For fuller details, the reader is referred to the previous paper (Ruben, 1961).

THE NON-EXISTENCE OF SOME PARTIALLY BALANCED INCOMPLETE BLOCK DESIGNS WITH LATIN SQUARE TYPE ASSOCIATION SCHEME

By S. S. SHRIKHANDE and N. C. JAIN

Banaras Hindu University

SUMMARY. A method for proving the impossibility of Symmetrical Partially Balanced Incomplete Block (PBIB) design with L_i association scheme is obtained. The technique employed is that used by Ogawa (1959) in proving the nonexistence of some PBIB designs of the triangular type.

1. INTRODUCTION AND SOME PRELIMINARY RESULTS

Two symmetric and nonsingular matrices A and B of the same order n with rational elements are said to be rationally congruent, if there exists a rational nonsingular matrix C of the same order such that

$$C'AC = B \qquad \ldots (1.1)$$

where C' is the transpose of C (Jones, 1950). We denote this relation by the notation

$$A \sim B. \qquad \ldots (1.2)$$

It is obvious that this relation satisfies the usual properties of an equivalence relation.

Denote the n leading principal minor determinants of A by $D_1, D_2, \ldots, D_n = |A|$ and let $D_0 = 1$, then the Hasse-Minkowski p-invariant of A is given by

$$C_p(A) = (-1, -1)_p \prod_{i=0}^{n-1} (D_{i+1}, -D_i)_p \qquad \ldots (1.3)$$

for each prime p, where $(a, b)_p$ denotes the extended Hilbert norm residue symbol (Jones, 1950; Hensel, 1915) defined by

$$(a, b)_p = \begin{cases} 1 & \text{if } ax^2 + by^2 = 1 \text{ has a } p\text{-adic solution} \\ -1 & \text{otherwise.} \end{cases} \qquad \ldots (1.4)$$

Then from Hasse (1923) and Jones (1950) we have

Hasse's Theorem : *The necessary and sufficient conditions for two positive definite, rational and symmetric matrices A and B of the same order to be rationally congruent are that the square-free parts of the determinants of both the matrices are the same and further the Hasse-Minkowski p-invariants of both the matrices are equal for all primes p including p_∞.*

259

We state without proof certain lemmas given by Ogawa (1959).

Lemma 1 : *If A, B and C are rational nonsingular and symmetric and if*

$$U = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}$$

$$V = \begin{pmatrix} A & 0 & 0 \\ 0 & B & 0 \\ 0 & 0 & C \end{pmatrix}$$

then $\quad c_p(U) = (-1, -1)_p \, (|A|, \, |B|)_p \, c_p(A) \, c_p(B).$

$c_p(V) = (|A|, \, |B|)_p \, (|A|, \, |C|)_p \, (|B|, \, |C|)_p \, c_p(A) \, c_p(B) c_p(C).$

Lemma 2 : *For an $n \times n$ diagonal matrix Δ_n, with each diagonal element d*

$$c_p(\Delta_n) = (-1, -1)_p \, (-1, d)_p^{\frac{n(n+1)}{2}} .$$

Lemma 3 : $c_p(\rho A) = (-1, \rho)_p^{\frac{n(n+1)}{2}} \, (\rho, \, |A|)_p^{n-1} \, c_p(A)$ *where n is the order of A.*

Lemma 4 : *If the $n-1$ rational column vectors*

$$a_2, a_3 \ldots, a_n$$

of dimensionality n are linearly independent and are orthogonal to

$$1' = (1, 1, \ldots\ldots 1)$$

then for the Gramian of the set, i.e.

$$U = \begin{pmatrix} a_2' \\ \vdots \\ a_n' \end{pmatrix} (a_2 \ldots a_n)$$

$$c_p(U) = (-1, -1)_p.$$

Lemma 5 : *So long as we restrict ourselves to rational vectors, the p-invariant of the Gramian of the set is uniquely determined by the linear subspace generated by the set.*

Lemma 6 : *If*

$$A = eI_n + fG_n$$

where I_n is the unit matrix of order n and G_n is the square matrix of order n with all elements unity, then

$$|A| = ge^{n-1}$$

where $\quad g = e + nf.$

If A and B are square matrices of order n, and X is a square matrix of order ns of the form

$$X = \begin{pmatrix} A & B & \ldots & B \\ B & A & \ldots & B \\ \vdots & \vdots & & \vdots \\ B & B & \ldots & A \end{pmatrix}$$

then $\quad |X| = |A - B|^{s-1} |A + (s-1) \, B|.$

We quote for completeness the properties of Hilbert symbol and some of the useful properties of the Legendre symbol (a/p) where p is a prime (Ogawa, 1959).

Lemma 7 :

$$(a, b)_p = (b, a)_p.$$

$$(at^2, bu^2)_p = (a, b)_p$$

if t and u are any rational numbers. We use the notation $a_1 \sim a_2$ to indicate that a_1 and a_2 differ only by a factor which is the square of a rational number. Hence in the calculation of the Hilbert symbol the square part of any rational number can be replaced by 1.

$$(a, -a)_p = 1.$$

$$(a, a)_p = (-1, a)_p.$$

$$(a, b_1 b_2)_p = (a, b_1)_p (a, b_2)_p.$$

$$(a, b)_p = (-ab, a+b)_p.$$

As a special case of the above we have for any positive integer n

$$(n, n+1)_p = (-1, n+1)_p.$$

For any odd prime p $\qquad (a, p)_p = (a/p)$

and if integers a and b are not divisible by p, then

$$(a, b)_p = 1.$$

For the Legendre symbol we have

$$(a/p) = (b/p) \quad \text{if} \quad a \equiv b (\bmod\ p),$$

$$(ab/p) = (a/p)\ (b/p),$$

$$(p/q)(q/p) = (-1)^{\frac{(p-1)}{2}\frac{(q-1)}{2}},$$

$$(-1/p) = (-1)^{\frac{p-1}{2}},\ (2/p) = (-1)^{\frac{p^2-1}{8}},$$

where p and q denote odd primes.

Let M be a positive definite, rational, symmetric and stochastic matrix of order v. Let ρ_0 be the sum of the elements in any row. Then ρ_0 is an eigen-value of M with multiplicity $n_0 = 1$. Suppose further that M has two other distinct and rational eigen-values ρ_1 and ρ_2 with multiplicities n_1 and n_2 where $n_0 + n_1 + n_2 = v$. Let X_i be a matrix of order (v, n_i) such that its columns generate the eigen-space of M corresponding to ρ_i; $i = 0, 1, 2$. We can assume without loss of generality that

$$X_i' X_i = I_{n_i} \qquad \qquad \dots \ (1.5)$$

$$X_i' X_j = 0_{n_i, n_j} \qquad \qquad \dots \ (1.6)$$

where any column of X_i consists of rational numbers divided by the square-root of the same positive number which depends upon the column number. In particular X_0 is a column vector with each component $1/\sqrt{v}$. Then it is easy to show that

$$M = \sum_{0}^{2} \rho_i A_i \qquad \qquad \dots \ (1.7)$$

where $\qquad\qquad\qquad A_i = (a_j^i) = X_i X_i' \qquad\qquad \dots \ (1.8)$

is a rational symmetric matrix of order v. In particular A_0 is a matrix with all elements $1/v$. Further from (1.5) and (1.6)

$$A_i^2 = A_i \qquad \qquad \text{...} \quad (1.9)$$

$$A_i A_j = 0_{v,v}. \qquad \qquad \text{...} \quad (1.10)$$

From (1.7), (1.9), (1.10) it follows that

$$MA_i = \rho_i A_i \qquad \qquad \text{...} \quad (1.11)$$

which means that A_i is the eigen-space corresponding to ρ_i. Without loss of generality assume that

$$a^0, a_1^1, a_2^1, \ldots a_{n_1}^1, a_1^2, a_2^2 \ldots a_{n_2}^2$$

are independent column vectors and put

$$S = \left(a^0, a_1^1 \ldots a_{n_1}^1 \; a_1^2 \ldots a_{n_2}^2 \right) = (B_0 \; B_1 \; B_2). \qquad \text{...} \quad (1.12)$$

Then S is a non-singular matrix with rational elements. Put

$$Q_i = B_i' B_i, \quad i = 0, 1, 2. \qquad \qquad \text{...} \quad (1.13)$$

Then
$$S' MS = \begin{pmatrix} \rho_0/v & 0 & 0 \\ 0 & \rho_1 Q_1 & 0 \\ 0 & 0 & \rho_2 Q_2 \end{pmatrix} \qquad \text{...} \quad (1.14)$$

or
$$M \sim \begin{pmatrix} \rho_0 v & 0 & 0 \\ 0 & \rho_1 Q_1 & 0 \\ 0 & 0 & \rho_2 Q_2 \end{pmatrix} \qquad \text{...} \quad (1.15)$$

Since
$$S'S = \begin{pmatrix} 1/v & 0 & 0 \\ 0 & Q_1 & 0 \\ 0 & 0 & Q_2 \end{pmatrix}$$

We obviously have

$$|Q_1| |Q_2| \sim v, \qquad \qquad \text{...} \quad (1.16)$$

Also, the columns of $(B_1 \; B_2)$ are orthogonal to the vector 1. Hence from Lemmas 1 and 4 and the fact that the Gramian of $(B_1 \; B_2)$ is

$$Q = \begin{pmatrix} Q_1 & 0 \\ 0 & Q_2 \end{pmatrix}$$

We have

$$c_p(Q) = (-1, -1)_p \; (|Q_1|, \; |Q_2|)_p \; c_p(Q_1) \; c_p(Q_2) = (-1, -1)_p$$

or
$$(|Q_1|, \; |Q_2|)_p \; c_p(Q_1) \; c_p(Q_2) = 1. \qquad \qquad \text{...} \quad (1.17)$$

From (1.15) to (1.17) using Lemmas 1 and 3 and the properties of the norm-residue symbols we have

$$c_p(M) = (-1, -1)_p \, (\rho_0, -v\rho_1^{n_1}\rho_2^{n_2})_p \, (\rho_2, v)_p \, (\rho_1 \rho_2, |Q_1|)_p$$

$$(\rho_1, \rho_2)_p^{n_1 n_2} \, (-1, \rho_1)_p^{\frac{n_1(n_1+1)}{2}} \, (-1, \rho_2)_p^{\frac{n_2(n_2+1)}{2}} \qquad \ldots \quad (1.18)$$

Hence we can state the following theorem.

Theorem 1 : *If M is a positive definite, rational, symmetric and stochastic matrix of order v with rational eigen-values ρ_0, ρ_1, ρ_2 with multiplicities $n_0(=1), n_1, n_2$ and Q_1 is the Gramian of the rational vectors generating the eigen-space corresponding to ρ_1, then $c_p(M)$ is given by the formula* (1.18).

L_i association scheme is defined by Bose and Shimamoto (1952) as follows : The number of treatments is n^2 where n is a positive integer. We assume that $i-2$ mutually orthogonal Latin Squares (MOLS) of order n exist (Bose, 1938). We arrange the n^2 treatments arbitrarily in a $n \times n$ square L. Let each of the $i-2$ MOLS L_j be superimposed on L. For each treatment θ in L, define its 1-associates as those treatments which lie in the same row or column of L as θ or those which correspond to the same symbol of L_j as θ when L_j is superimposed on $L; j = 1, 2, ..., i-2$. The remaining treatments are called 2-associates. In this scheme each treatment has n_i i-associates $i = 1, 2$, where

$$n_1 = i(n-1), \quad n_2 = (n+1-i)\,(n-1). \qquad \ldots \quad (1.19)$$

The parameters of the second kind are given by

$$P_1 = \begin{pmatrix} (n-2)+(i-1)(i-2) & (i-1)(n+1-i) \\ & (n-i)(n+1-i) \end{pmatrix} \qquad \ldots \quad (1.20)$$

$$P_2 = \begin{pmatrix} i(i-1) & i(n-i) \\ & (n-2)+(n-i)(n-i-1) \end{pmatrix}.$$

A symmetrical PBIB design with L_i scheme is an arrangement of $v = n^2$ treatments in v block of size $r < v$ satisfying the following conditions :

(i)   Each treatment occurs in exactly r blocks.

(ii) No treatment occurs more than once in any block.

(iii) Any two treatments which are i-associates as defined above occur together in exactly λ_i blocks, $i = 1, 2$.

Consider a matrix with i rows and n^2 columns in n distinct symbols having the property that in any two rows every ordered pair $\binom{a}{b}$ occurs exactly once. Such a scheme is called an orthogonal array of strength 2, index unity and i constraints (Rao, 1947) and is denoted by $(n^2, i, n, 2)$. We state an alternative definition of L_i scheme as a lemma.

263

Lemma 8 : *Suppose an orthogonal array* $(n^2, i, n, 2)$ *exists. Identify the columns with n^2 treatments in any arbitrary manner. Define 1-associates of any treatment θ as those treatments for which the corresponding columns coincide with the column corresponding to θ in exactly one position. Define the remaining treatments as 2-associates of θ. Then the association scheme thus defined is of L_i type.*

The proof of this lemma follows from the equivalence of the existence of $i-2$ MOLS of order n and the orthogonal array $(n^2, i, n, 2)$ as shown by Rao (1947).

2. Main result

Suppose a symmetrical PBIB design with L_i scheme for $v = n^2$ treatments with parameters r, λ_1, λ_2 exists. Define the incidence matrix $N = (n_{ij})$ in the usual manner.

$$n_{ij} = \begin{cases} 1, \text{ if treatment } i \text{ occurs in block } j, \\ \\ 0 \text{ otherwise.} \end{cases}$$

Then $$M = NN' \qquad \ldots \ (2.1)$$

is a nonnegative rational, symmetric and stochastic matrix with eigen-values (Connor and Clatworthy, 1954).

$$\begin{aligned} \rho_0 &= r^2 \\ \rho_1 &= r + (n-i)\,\lambda_1 - (n+1-i)\lambda_2 \qquad \ldots \ (2.2) \\ \rho_2 &= r - i\lambda_1 + (i-1)\lambda_2 \end{aligned}$$

with multiplicities $\qquad n_0 = 1, n_1 \text{ and } n_2,$

where n_1 and n_2 are given by (1.19). We consider the case when ρ_1 and ρ_2 are both positive in which case M satisfies all the conditions of Theorem 1. Then from

$$|M| = r^2 \ \rho_1^{n_1} \ \rho_2^{n_2} = |N|^2 \qquad \ldots \ (2.3)$$

it follows that when n is even then ρ_1 must be a perfect square if i is odd and ρ_2 must be a perfect square if i is even.

We use a method similar to that of Corsten (1960) to obtain the eigen-space corresponding to the value ρ_1 of M. Let $(n^2, i, n, 2)$ be an orthogonal array in n integers $1, 2, \ldots, n$. Number the columns from 1 to n^2 and identify the treatments with the columns. Let c_{kj} be a column vector with v components formed in the following manner. Suppose the integer j occurs in the k-th row of the orthogonal array in columns numbered $k_{j1}, \ldots k_{jn}$, then c_{kj} contains 1 in exactly these positions and 0 elsewhere, $k = 1, 2, \ldots, i; \ j = 1, 2, \ldots, n$. From the properties of the orthogonal array it is easy to see that for every k

$$\sum_{j=1}^{n} c_{kj} = 1 \qquad \ldots \ (2.4)$$

and that the set of ni vectors thus formed is of rank $i(n-1)+1$. Denote the vector space generated by these vectors by B. Then B can also be generated by $\mathbf{1}$ and $n-1$ vectors selected from each of the i sets $C_k = (c_{kj}), j = 1, 2, \ldots, n$, which form a rational

basis of B. Let B_1 be the subspace of B which is orthogonal to 1. Let $x = (x_\alpha)$, $\alpha = 1, 2, \ldots, n^2$, be any vector of B_1, then

$$x = \sum_1^i \sum_1^n e_{kj} \, c_{kj} \qquad \qquad \ldots \; (2.5)$$

where

$$\Sigma\Sigma \, e_{kj} = 0. \qquad \qquad \ldots \; (2.6)$$

The sum of the coordinates of x corresponding to all the treatments which are 1-associates of α (including $x_\alpha i$ times) is

$$i x_\alpha + S_1 \qquad \qquad \ldots \; (2.7)$$

where S_1 stands for the sum of the coordinates whose positions correspond to 1-associates of α. If α corresponds to the column

$$\begin{pmatrix} j_1 \\ j_2 \\ \cdot \\ \cdot \\ j_i \end{pmatrix}$$

of $(n^2, i, n, 2)$, then from Lemma 8 we have

$$x_\alpha = \sum_{k=1}^i e_{kj_k} \qquad \qquad \ldots \; (2.8)$$

$$S_1 = (n-1)\sum_{k=1}^i e_{kj_k} + \sum_{m=1}^i b_m \qquad \qquad \ldots \; (2.9)$$

where

$$b_m = \sum_{k=1}^i a_k - a_m \qquad \qquad \ldots \; (2.10)$$

and

$$a_k = \sum_{j=1}^n e_{kj} - e_{kj_k}.$$

Using (2.6) it is easy to verify that

$$i x_\alpha + S_1 = n x_\alpha$$

$$S_1 = (n-i) x_\alpha \qquad \qquad \ldots \; (2.11)$$

Hence

for all vectors x in B_1. If S_2 denotes the sum of the coordinates of x corresponding to 2-associates of α, then from the fact that x is orthogonal to 1

$$x_\alpha + S_1 + S_2 = 0. \qquad \qquad \ldots \; (2.12)$$

Putting

$$y = Mx$$

for all x in B_1 we have

$$y_\alpha = r x_\alpha + \lambda_1 S_1 + \lambda_2 S_2 \qquad \qquad \ldots \; (2.13)$$

and hence from (2.11) and (2.12) $y_\alpha = \rho_1 x_\alpha$.

Thus any vector of B_1 is an eigen-vector of M corresponding to ρ_1 and since the rank of B_1 is $n_1 = i(n-1)$ this implies that the eigen-space corresponding to ρ_1 is exactly B_1.

Since B is generated by the rational vectors 1 and $n-1$ vectors selected from each of the i sets C_k, as also by 1 and a basis of B_1 which can be taken as rational, the two expressions for the Gramian are rationally congruent. Denote by Q_1 the Gramian of B_1. Noting that

$$1'1 = n^2,$$

$$1'c_{kj} = n,$$

$$c'_{kj}\, c_{kj} = n,$$

$$c'_{kj}\, c_{kj'} = 0 \quad \text{if } j \neq j',$$

$$c'_{kj}\, c_{k'j'} = 1 \quad \text{if } k \neq k'.$$

We obviously have

$$
\begin{pmatrix} n^2 & 0 \\ 0 & Q_1 \end{pmatrix} \sim
\begin{pmatrix}
n^2 & nJ' & nJ' & \cdots & nJ' \\
nJ & nI_{n-1} & G_{n-1} & \cdots & G_{n-1} \\
nJ & G_{n-1} & nI_{n-1} & \cdots & G_{n-1} \\
\cdot & \cdot & \cdot & & \cdot \\
\cdot & \cdot & \cdot & & \cdot \\
nJ & G_{n-1} & G_{n-1} & \cdots & nI_{n-1}
\end{pmatrix}
$$

where J is a column vector with $n-1$ components all unities. The determinant of the matrix on the right hand side can be evaluated by performing elementary transformations on the rows and the same transformations on the columns and making use of Lemma 6. It is then easily verified that

$$|Q_1| \sim n^{in}. \qquad \ldots (2.14)$$

Substituting in (1.18) and using the properties of the Hilbert symbol, we have for any odd prime p

$$c_p(M) = (\rho_1, \rho_2)_p^{n_1 n_2} \; (-1, \rho_1)_p^{\frac{n_1(n_1+1)}{2}} \; (-1, \rho_2)_p^{\frac{n_2(n_2+1)}{2}} \; (\rho_1 \rho_2, n)_p^{in} \qquad \ldots (2.15)$$

where n_1, n_2, ρ_1 and ρ_2 are given by (1.19) and (2.2). Hereafter we omit to write p in the symbol $(a, b)_p$. We consider the two cases (i) i odd, (ii) i even.

Case (i), *i odd* : If n is even we already have a necessary condition that ρ_1 be a perfect square. Then

$$c_p(M) = (-1, \rho_2)^{\frac{n_2(n_2+1)}{2}}$$

The index of the expression on the right hand side is odd if and only if $n+1 \equiv 3$ (mod 4), in which case

$$c_p(M) = (-1, \rho_2) = -1$$

if and only if the square-free part of ρ_2 contains a prime $\equiv 3$ (mod 4).

If n is odd then

$$c_p(M) = (-1, \rho_1)^{\frac{n_1(n_1+1)}{2}} (-1, \rho_2)^{\frac{n_2(n_2+1)}{2}} \cdot (\rho_1 \rho_2, n).$$

It is easy to verify that in this case the above gives

$$c_p(M) = \left((-1)^{\frac{n-1}{2}} n, \rho_1 \rho_2 \right)$$

Case (ii), i *even :* Expression (2.15) reduces to

$$c_p(M) = (-1, \rho_1)^{\frac{n_1(n_1+1)}{2}} (-1, \rho_2)^{\frac{n_2(n_2+1)}{2}}.$$

If n is even a necessary condition for existence of N is that ρ_2 be a perfect square, in which case

$$c_p(M) = (-1, \rho_1)^{\frac{n_1(n_1+1)}{2}}.$$

The index in the right hand expression is odd if and only if $i \equiv 2 \pmod 4$ in which case

$$c_p(M) = (-1, \rho_1)$$
$$= -1.$$

if and only if the square-free part of ρ_1 contains a prime $\equiv 3 \pmod 4$.

Noting that $\qquad M \sim I_v,$

we have $\qquad c_p(M) = c_p(I_v) = 1.$

Hence we can state the following two theorems.

Theorem 2 : *Necessary conditions for the existence of a symmetrical PBIB design for n^2 treatments with L_i association scheme when ρ_1 and ρ_2 are positive and i is odd are :*

(i) *if n is even then ρ_1 must be a perfect square and if further $n+i \equiv 3 \pmod 4$, then the square-free part of ρ_2 does not contain a prime $\equiv 3 \pmod 4$; and if*

(ii) *n is odd then* $\left((-1)^{\frac{n-1}{2}} n, \rho_1 \rho_2 \right)_p = 1$, *where p is an odd prime.*

Theorem 3 : *Necessary condition for the existence of a symmetrical PBIB design for n^2 treatments with L_i association scheme when ρ_1 and ρ_2 are positive and i is even is that if n is even ρ_2 must be a perfect square and if further $i \equiv 2 \pmod 4$, then the square-free part of ρ_1 does not contain a prime $\equiv 3 \pmod 4$.*

It is interesting to note that Theorem 3 is obtained from (i) of Theorem 2, by replacing i by $n+1-i$ and interchanging λ_1 and λ_2 which implies interchanging ρ_1 and ρ_2.

We indicate in the following table the nonexistence of certain designs. The last column indicates the application of the proper theorem together with value of p when necessary.

TABLE. IMPOSSIBILITY OF CERTAIN DESIGNS

n	i	r	λ_1	λ_2	remarks
5	3	4	0	1	Theorem 2(ii), p=3
5	3	4	1	0	,, ,, ,,
7	3	15	5	4	,, ,, ,,
7	5	15	4	5	,, ,, ,,
7	3	21	10	8	,, ,, ,,
7	5	21	8	10	,, ,, ,,
7	3	24	14	10	,, ,, ,,
7	5	24	10	14	,, ,, ,,
10	5	18	2	4	Theorem 2 (i)
10	6	18	4	2	Theorem 3
12	6	45	16	12	Theorem 3
12	7	45	12	16	Theorem 2(i)
14	5	40	6	9	Theorem 2 (i)
14	10	40	9	6	Theorem 3
14	6	27	6	2	Theorem 3
14	9	27	2	6	Theorem 2(i)

In conclusion we note that Theorem 1 can be generalised to prove the impossibility of m associate classes symmetrical PBIB design whenever the eigen-values of $M = NN'$ are rational and positive and it is possible to find the Gramians of the bases of $m-1$ of the m eigen-spaces corresponding to the values $\rho_1, \rho_2, ..., \rho_m$. We also note that in case of two associate classes symmetrical PBIB design if $\rho_1 \sim \rho_2$, then one need not find the value of $|Q_1|$ at all. These points will be considered in a later communication.

References

BOSE, R. C. (1938) : On the application of the properties of Galois fields to the construction of Hyper-Graeco Latin Squares. *Sankhyā*, **3**, 323-338.

BOSE, R. C., and SHIMAMOTO, T. (1952) : Classification and analysis of partially balanced incomplete block designs with two associate classes. *J. Amer. Stat. Ass.*, **47**, 151-184.

CONNOR, W. S. and CLATWORTHY, W. H. (1954) : Some theorems for partially balanced designs. *Ann. Math. Stat.*, **28**, 100-112.

CORSTEN, L. C. A. (1960) : Proper spaces related to triangular PBIB designs. *Ann. Math. Stat.*, **31**, 498-502.

HASSE, H. (1923) : Uber die Aquivalenz quadratische Formen im Korper der rationalen Zahlen. *J. Reine Angew. Math.*, **152**, 205-224.

HENSEL, K. (1915) : Zahlentheorie, G. J. Göschen' sche Verlagschandlung G.m.b.H., Berlin and Leipzig.

JONES, B. W. (1950) : *The Arithmetic Theory of Quadratic Forms*, The Mathematical Association of America, John Wiley and Sons, Inc.

OGAWA, J. (1959) : A Necessary condition for existence of regular and symmetrical experimental designs of triangular type, with partially balanced incomplete blocks. *Ann. Math. Stat.*, **30**, 1063-1071.

RAO, C. R. (1947) : Factorial experiments derivable from combinatorial arrangements of arrays. *J. Roy. Stat. Soc. Suppl.*, **9**, 128-139.

Paper received : December, 1961.

RECOVERY OF INTERBLOCK INFORMATION

By J. ROY and K. R. SHAH

Indian Statistical Institute

SUMMARY. The problem of recovery of inter-block information for a general incomplete block design is examined in this paper. The ratio ρ of the inter-block variance to the intra-block variance plays a key role. Under the so-called Normal set up, the usual estimator of ρ is biased; expressions for the bias and variance are derived. Several alternative estimators of ρ having desirable properties are examined. A computational procedure for obtaining the maximum likelihood estimate is given. If a certain type of estimator of ρ is used, the estimators of treatment effects are proved to be unbiased. An expression is derived for the increase in the variance of the estimate of a treatment effect due to fluctuations of sampling in ρ.

1. INTRODUCTION

Since some incomplete block designs have low efficiency factors, Yates suggested the use of information available from inter-block comparisons to increase the precision of estimates of treatment effects. He called the process recovery of inter-block information and showed how this is to be done for a cubic lattice design (1939) and a balanced incomplete block (BIB) design (1940). Nair (1944) gave the method for partially balanced incomplete block designs and finally Rao (1947, 1956) adopted the method for any incomplete block design.

The process consists in applying the method of weighted least squares to intra-block contrasts and inter-block contrasts of observations, for the purpose of estimating the treatment effects, weights being inversely proportional to the variances of these contrasts. The ratio ρ of the inter-block variance to the intra-block variance plays a key role and since this is usually unknown, the ratio of estimates of these variances obtained from an analysis of variance of the data is substituted. The properties of estimates so obtained have not so far been investigated in detail. Recently, Graybill and Weeks (1959) have proved that under the so-called normal model, estimates of treatment effects so obtained, are unbiased in the case of BIB designs and they have also obtained (1961) a minimal set of sufficient statistics.

The problem of recovery of inter-block information in the case of a general incomplete block design is examined in detail in this paper. The usual estimator of the variance ratio ρ obtained from the analysis of variance is found to be biased and a simple correction is obtained for this bias. An expression for the variance of this estimator is derived. Some alternative estimators of ρ having desirable properties and based on quadratic estimators of inter-block and intra-block variances are also proposed. The method of maximum likelihood is shown to give rise to a somewhat complicated equation for estimation, and a numerical procedure for solving the equation by iteration is presented.

It is shown further that if a certain type of estimator of the variance ratio is used, the final estimators of treatment effects turn out to be unbiased. It is shown that there is an increase in the sampling error of the treatment effects due to the sampling fluctuation in the variance ratio, and an expression for this increment is derived.

Consider an experiment in which v treatments are applied on bk experimental units or plots, themselves divided into b blocks of k plots each. Only one treatment is applied on every plot, the actual allocation being done in the following manner.

First we consider a design, that is an arrangement of v symbols (one corresponding to each treatment) in b rows, each having k cells. The arrangement is characterised by the numbers m_{jiu}, $j = 1, 2, ..., v$; $i = 1, 2, ..., b$; $u = 1, 2, ..., k$ where $m_{jiu} = 1$ or 0 according as the j-th symbol (treatment) occurs on the u-th cell of the i-th row or not.

Next, the blocks are numbered $1, 2, ..., b$ at random and the plots in a block are numbered $1, 2, ... k$ again at random and independently for different blocks. The u-th plot in the i-th block then receives the treatment corresponding to the symbol which occurs in the u-th cell of the i-th row of the design. Let y_{iu} denote the yield of this plot. Under the assumption that the yield from any plot is the sum of two components, one due to the plot and the other due to the treatment, we have

$$E(y_{iu}) = \mu + \sum_{j=1}^{v} \theta_j m_{jiu} \qquad \ldots \ (1.1)$$

$$\text{cor} (y_{iu}, y_{i'u'}) = \begin{cases} \left(1 - \dfrac{1}{k}\right)\sigma_0^2 + \dfrac{1}{k}\left(1 - \dfrac{1}{b}\right)\sigma_1^2 & \text{if } i \doteq i', \ u = u' \\[2mm] -\dfrac{1}{k}\,\sigma_0^2 + \dfrac{1}{k}\left(1 - \dfrac{1}{b}\right)\sigma_1^2 & \text{if } i = i', \ u \neq u' \qquad \ldots \ (1.2) \\[2mm] -\dfrac{1}{bk}\,\sigma_1^2 & \text{if } i \neq i'. \end{cases}$$

Here θ_j is the effect of the j-th treatment $\sum_{j=1}^{v} \theta_j = 0$, μ is the overall mean of plot effects, and σ_0^2 and σ_1^2 are respectively the mean squares (of plot effects) within and between the blocks. From Section 3 onwards, we shall make the further assumption that the joint distribution of the random variables y_{iu}'s is multivariate normal, with first and second order moments given by (1.1) and (1.2). We shall write

$$\rho = \sigma_1^2/\sigma_0^2. \qquad \ldots \ (1.3)$$

The problem is to estimate the parameters θ_j's and σ_0^2 and σ_1^2.

Let $\sum_{u=1}^{k} m_{jiu} = n_{ji}$, the number of times the j-th treatment occurs on plots in the i-th block. Thus $n_{ji} = 1$ or 0 and $\sum_{j=1}^{v} n_{ji} = k$, $\sum_{i=1}^{b} n_{ji} = r$. The $v \times b$ matrix $N = ((n_{ji}))$ is called the *incidence matrix*.

We shall denote by $E_{m \times n}$ a matrix of the form $m \times n$, each element of which is unity. The matrices

$$C = rI - \frac{1}{k} \ NN' \quad \text{and} \quad C_1 = \frac{1}{k} \ NN' - \frac{r^2}{bk} \ E_{vv} \qquad \ldots \ (1.4)$$

play important roles in the analysis. We shall assume that the matrix C is of rank $(v-1)$: this is equivalent to the assumption that the experimental design is *connected*.

A linear function $\sum_{i,u} a_{iu} y_{iu}$ is said to be a *contrast* if $\sum_{i,u} a_{iu} = 0$. A contrast is said to belong to blocks, or simply called an *inter-block contrast* if $a_{i1} = a_{i2} = \cdots = a_{ik}$ holds for all i. A contrast is said to be an *intra-block contrast* if $\sum_{u} a_{iu} = 0$ holds for all i. A linear function $\sum_{i,u} a_{iu} y_{iu}$ is said to be *normalised* if $\sum_{i,u} a_{iu}^2 = 1$. Two linear functions $\sum_{i,u} a_{iu} y_{iu}$ and $\sum_{i,u} b_{iu} y_{iu}$ are said to be *orthogonal* if $\sum_{i,u} a_{iu} b_{iu} = 0$. It is easy to see that any inter-block contrast and any intra-block contrast are mutually orthogonal. The rank of the vector-space generated by all inter-block contrasts is $(b-1)$ and of that generated by all intra-block contrasts is $b(k-1)$.

Let B_i denote the total yield for the i-th block and T_j that for the j-th treatment and let G be the grand total, so that

$$B_i = \sum_{u} y_{iu}, \quad T_j = \sum_{i,u} y_{iu} m_{jiu} \quad \text{and} \quad G = \sum_{i,u} y_{iu}. \qquad \ldots \text{(1.5)}$$

We shall use the row-vectors $B = (B_1, B_2, \ldots, B_b)$, and $T = (T_1, T_2, \ldots, T_v)$. The adjusted yields for the treatments are defined as

$$Q = T - \frac{1}{k} BN'. \qquad \ldots \text{(1.6)}$$

Let, further

$$Q_1 = \frac{1}{k} BN' - \frac{rG}{bk} E_{1v}. \qquad \ldots \text{(1.7)}$$

It can be seen that the elements of Q are intra-block contrasts and those of Q_1 are inter-block contrasts.

It is known (see, for example, Rao, 1947) that minimum variance unbiased linear estimates of the treatment effects $\theta = (\theta_1, \theta_2, \ldots, \theta_v)$, based on intra-block contrasts only, are obtained from the equations

$$\theta C = Q. \qquad \ldots \text{(1.8)}$$

We shall write θ^* for the solution of these equations. If the ratio $\rho = \sigma_1^2 / \sigma_0^2$ is known, both intra-block and inter-block contrasts can be used together, and minimum variance linear unbiased estimates in this case are obtained from the equation

$$\theta \left(C + \frac{1}{\rho} C_1 \right) = Q + \frac{1}{\rho} Q_1. \qquad \ldots \text{(1.9)}$$

The solution of these equations will be denoted by $\bar{\theta}(\rho)$. When ρ is not known, an estimate ρ^* for ρ is substituted in (1.9) and $\bar{\theta}(\rho^*)$ is taken as an estimate for θ.

9

For estimating ρ, the following procedure is generally recommended. (See Yates, 1939, 1940 or, for a general treatment, Rao, 1947). First the following table of analysis of variance is prepared :

TABLE   ANALYSIS OF VARIANCE

source	d.f.	s.s.
blocks (unadjusted)	$b-1$	$SS_B^* = \frac{1}{k}\mathbf{BB}' - G^2/bk$
treatments (adjusted)	$v-1$	$SS_{tr} = \mathbf{Q\theta}^{*\prime}$
error	$e_0 = bk - b - v + 1$	$SS_E = SS_T - SS_B^* - SS_{tr}$
total	$bk-1$	$SS_T = \sum\limits_{i,u} y_{iu}^2 - G^2/bk$

The adjusted sum of squares due to blocks is then computed as

$$SS_B = SS_B^* + SS_{tr} - \left(\frac{1}{r}\mathbf{TT}' - G^2/bk\right). \qquad \ldots \quad (1.10)$$

Then s_0^2 and s_1^2 defined by

$$s_0^2 = SS_E/e_0, \quad v(r-1)s_1^2 = kSS_B - (v-k)s_0^2 \qquad \ldots \quad (1.11)$$

provide unbiased estimators of σ_0^2 and σ_1^2 respectively; and as an estimate of ρ one takes

$$R = s_1^2/s_0^2. \qquad \ldots \quad (1.12)$$

If the blocks are formed so as to achieve homogeneity within blocks, we expect $\rho \geqslant 1$; but depending on fluctuations of sampling, R may not satisfy this inequality. For this reason, a modified estimate R' (which we shall call the *truncated* form of R) given by

$$R' = \begin{cases} 1 & \text{if } R \leqslant 1 \\ R & \text{if } R \geqslant 1 \end{cases} \qquad \ldots \quad (1.13)$$

has at times been recommended.

2.   CANONICAL REDUCTION

The assumption that the rank of C is $(v-1)$ implies that there is exactly one latent root of the matrix NN' which is equal to rk, and all other latent roots are strictly smaller than rk. Let ξ_s, $s = 1, 2, \ldots, q$ be a set of orthonormal latent vectors of NN', corresponding to the q positive latent roots ϕ_s, all smaller than rk. Let ξ_s, $s = q+1$, $q+2, \ldots, v-1$ be a set of $(v-1)-q$ orthonormal $1 \times v$ vectors, each orthogonal to $\xi_1, \xi_2, \ldots, \xi_q$ and also to E_{1v}. We then define $(v-1)$ intra-block contrasts x_{0s}; $s = 1, 2, \ldots, v-1$ as follows :

$$x_{0s} = \begin{cases} k^{\frac{1}{2}}(rk-\phi_s)^{-\frac{1}{2}}\mathbf{Q}\xi_s' & \text{for } s = 1, 2, \ldots, q \\ r^{-\frac{1}{2}}\mathbf{Q}\xi_s' & \text{for } s = q+1, q+2, \ldots, v-1. \end{cases} \qquad \ldots \quad (2.1)$$

Since the rank of the vector-space generated by all intra-block contrasts is $b(k-1)$, we can find $e_0 = b(k-1) - (v-1)$ mutually orthogonal normalised intra-block contrasts, call them z_{0s}, $s = 1, 2, \ldots, e_0$, each orthogonal to $x_{01}, x_{02}, \ldots, x_{0, v-1}$.

Next, we define q inter-block contrasts

$$x_{1s} = (k\phi_s)^{-\frac{1}{2}}BN'\xi_s'; \quad s = 1, 2, ..., q. \qquad \qquad ... \quad (2.2)$$

Since the rank of the vector-space generated by inter-block contrasts is $(b-1)$, we can find $e_1 = (b-1)-q$ mutually orthogonal normalised inter-block contrasts; call them z_{1s}; $s = 1, 2, ..., e_1$, each orthogonal to $x_{11}, x_{12}, ..., x_{1q}$. Finally, let

$$G^* = (bk)^{-\frac{1}{2}}G. \qquad \qquad ... \quad (2.3)$$

By straightforward algebra, one can easily verify the following results.

2.1.   The linear transformation from y_{iu}'s to G^*, $x_{0s}(s = 1, 2, ..., v-1)$, $x_{1s}(s = 1, 2, ..., q)$, z_{0s} $(s = 1, 2, ..., e_0)$ and z_{1s} $(s = 1, 2, ..., e_1)$ is normalised orthogonal.

2.2.   The transformed variables are all mutually uncorrelated and their expectations and variances are :

$$E(x_{0s}) = a_{0s}\tau_s, \quad \text{where} \quad a_{0s} = \begin{cases} (r-\phi_s/k)^{\frac{1}{2}} & \text{for} \quad s = 1, 2, ..., q \\ r^{\frac{1}{2}} & \text{for} \quad s = q+1, q+2, ..., v-1, \end{cases} \qquad ... \quad (2.4)$$

$$E(x_{1s}) = a_{1s}\tau_s, \quad \text{where} \quad a_{1s} = (\phi_s/k)^{\frac{1}{2}} \qquad \text{for} \quad s = 1, 2, ..., q, \qquad ... \quad (2.5)$$

where
$$\tau_s = \theta\xi_s'; \quad s = 1, 2, ..., \quad v-1. \qquad ... \quad (2.6)$$

$$E(z_{0s}) = 0, \quad \text{for} \quad s = 1, 2, ..., e_0, \qquad E(z_{1s}) = 0 \quad \text{for} \quad s = 1, 2, ..., e_0$$

$$E(G^*) = (bk)^{\frac{1}{2}}\mu \qquad \qquad ... \quad (2.7)$$

$$V(x_{0s}) = \sigma_0^2, \quad \text{for} \quad s = 1, 2, ..., v-1, \quad V(z_{0s}) = \sigma_0^2, \quad \text{for} \quad s = 1, 2, ..., e_0 \quad ... \quad (2.8)$$

$$V(x_{1s}) = \sigma_1^2, \quad \text{for} \quad s = 1, 2, ..., q, \qquad V(z_{1s}) = \sigma_1^2, \quad \text{for} \quad s = 1, 2, ..., e_1$$

$$V(G^*) = 0. \qquad \qquad ... \quad (2.9)$$

2.3.   The equations (1.8) are equivalent to $\tau_s = t_s$ where

$$t_s = x_{0s}/a_{0s} \quad \text{for} \quad s = 1, 2, ..., v-1. \qquad \qquad ... \quad (2.10)$$

2.4.   The equations (1.9) are equivalent to $\tau_s = \tilde{t}_s(\rho)$ where

$$\tilde{t}_s(\rho) = \begin{cases} (\rho a_{0s}x_{0s}+a_{1s}x_{1s})/(\rho a_{0s}^2+a_{1s}^2), & \text{for } s = 1, 2, ..., q \\ x_{0s}/a_{0s} & \text{for } s = q+1, q+2, ..., v-1. \end{cases} \qquad ... \quad (2.11)$$

2.5.   The error sum of squares in the Table may be expressed as

$$SS_E = \sum_{s=1}^{e_0} z_{0s}^2 = S_0, \quad \text{say.} \qquad \qquad ... \quad (2.12)$$

2.6.   The adjusted sum of squares due to blocks defined by (1.10) may be expressed as

$$SS_B = S_1 + \sum_{s=1}^{q} \phi_s x_s^2/rk \qquad \qquad ... \quad (2.13)$$

273

where

$$S_1 = \sum_{s=1}^{e_1} z_{1s}^2 \qquad \qquad \dots \; (2.14)$$

and

$$z_s = x_{0s} - a_{0s} x_{1s}/a_{1s} \quad \text{for} \quad s = 1, 2, \dots, q. \qquad \dots \; (2.15)$$

2.7. Let

$$w_{iu}(\theta) = y_{iu} - \sum_j m_{jiu}\theta_j \qquad \qquad \dots \; (2.16)$$

$$B_i(\theta) = \sum_u w_{iu}(\theta). \qquad \qquad \dots \; (2.17)$$

Then

$$S_1 + \sum_{s=1}^q (x_{1s} - a_{1s}\tau_s)^2 = \frac{1}{k}\sum_i B_i^2(\theta) - \frac{G^2}{bk} \qquad \dots \; (2.18)$$

and

$$S_0 + \sum_{s=1}^{v-1} (x_{0s} - a_{0s}\tau_s)^2 = \sum_{i,u} w_{iu}^2(\theta) - \frac{1}{k}\sum_i B_i^2(\theta)$$

$$= \sum_{i,u} y_{iu}^2 - 2\sum_j \theta_j T_j + r\sum_j \theta_j^2 - \frac{1}{k}\sum_i B_i^2(\theta). \; \dots \; (2.19)$$

2.8. If the joint distribution of y_{iu}'s is in addition multivariate normal, a minimal set of sufficient statistics for the parameters θ, σ_0^2, and σ_1^2 is provided by x_{0s}, $(s = 1, \dots, v-1)$, x_{1s} $(s = 1, 2, \dots, q)$, S_0 and S_1. If ρ is given, $\bar{t}_s(\rho)$ $(s = 1, 2, \dots, v-1)$ and V are complete sufficient, where

$$V = S_0 + \frac{S_1}{\rho} + \sum_{s=1}^q \frac{z_s^2}{1 + \rho a_{0s}^2/a_{1s}^2}. \qquad \dots \; (2.20)$$

2.9. When ρ is known $\bar{t}_s(\rho)$ as defined by (2.11) is the unbiased minimum variance estimator of τ_s and its variance is given by

$$V\,\bar{t}_s(\rho) = \begin{cases} \rho\sigma_0^2/(\rho a_{0s}^2 + a_{1s}^2) & \text{for } s = 1, 2, \dots, q \\ \sigma_0^2/a_{0s}^2, & \text{for } s = q+1, \dots, v-1. \end{cases} \qquad \dots \; (2.21)$$

3. Maximum Likelihood Estimates

Under the assumption that the joint distribution of the random variables y_{iu}'s is multivariate normal with first and second order moments given by (1.1) and (1.2) it follows that the likelihood function L is given by

$$\log_e L = \text{const} - \frac{1}{2}\left[(b-1) \log_e \sigma_1^2 + b(k-1) \log_e \sigma_0^2 + \frac{1}{\sigma_1^2}\left\{ \sum_{s=1}^q (x_{1s} - a_{1s}\tau_s)^2 + S_1 \right\} \right.$$

$$\left. + \frac{1}{\sigma_0^2}\left\{ \sum_{s=1}^{v-1} (x_{0s} - a_{0s}\tau_s)^2 + S_0 \right\} \right] \qquad \dots \; (3.1)$$

where S_0 and S_1 are defined by (2.12) and (2.14) respectively. In all subsequent sections of this paper, we shall assume the joint distribution to be multivariate normal.

The likelihood equations, obtained by equating to zero the partial derivatives of $\log_e L$ with respect to the parameters, turn out to be

$$\tau_s = \bar{t}_s(\rho) \qquad \text{for } s = 1, 2, \ldots, v-1 \qquad \ldots \text{ (3.2)}$$

where $\bar{t}_s(\rho)$ is defined by (2.11) and

$$b(k-1)\sigma_0^2 = S_0 + \sum_{s=1}^{v-1} (x_{0s} - a_{0s}\tau_s)^2 \qquad \ldots \text{ (3.3)}$$

$$(b-1)\sigma_1^2 = S_1 + \sum_{s=1}^{q} (x_{1s} - a_{1s}\tau_s)^2. \qquad \ldots \text{ (3.4)}$$

The diagonal elements of the information matrix are

$$I(\tau_s, \tau_s) = \begin{cases} a_{0s}^2 \sigma_0^{-2} + a_{1s}^2 \sigma_1^{-2}, & \text{for } s = 1, 2, \ldots, q \\ a_{0s}^2 \sigma_0^{-2}, & \text{for } s = q+1, q+2, \ldots, v-1 \end{cases} \quad \ldots \text{ (3.5)}$$

$$I(\sigma_0^2, \sigma_0^2) = \tfrac{1}{2}b(k-1)\sigma_0^{-4} \qquad \ldots \text{ (3.6)}$$

$$I(\sigma_1^2, \sigma_1^2) = \tfrac{1}{2}(b-1)\sigma_1^{-4} \qquad \ldots \text{ (3.7)}$$

and all non-diagonal elements vanish.

We thus see that the maximum likelihood estimate of τ_s is $\hat{\tau}_s = \bar{t}_s(\hat{\rho})$ where $\hat{\rho}$ is the maximum likelihood estimate of ρ. To compute $\hat{\rho}$ we note that it can be expressed as

$$\hat{\rho} = \frac{b(k-1)\left[S_1 + \sum_{s=1}^{q} (x_{1s} - a_{1s}\hat{\tau}_s)^2\right]}{(b-1)\left[S_0 + \sum_{s=1}^{v-1} (x_{0s} - a_{0s}\hat{\tau}_s)^2\right]}. \qquad \ldots \text{ (3.8)}$$

We therefore use an iterative procedure. Starting with some suitable approximation for $\hat{\tau}_s$, we obtain a first approximation for $\hat{\rho}$ using (3.8). This value of $\hat{\rho}$ is used to obtain improved values for $\hat{\tau}_s$, which, in turn, when used in (3.8) provides a second better approximation for $\hat{\rho}$. This iterative procedure is continued till one gets stable values for $\hat{\tau}_s$'s and $\hat{\rho}$.

In actual computation, we do not work with the transformed canonical variables, but make use of the result 2.7 in Section 2. The iteration formula then is

$$\rho^{(n)} = \frac{b(k-1)\left[\frac{1}{k}\sum_i B_i^2(\theta^{(n)}) - \frac{G^2}{bk}\right]}{(b-1)\left[\sum_{i,u} y_{iu}^2 - 2\sum_j \theta_j^{(n)}T_j + r\sum_j [\theta_j^{(n)}]^2 - \frac{1}{k}\sum_i B_i^2(\theta^{(n)})\right]} \qquad \ldots \text{ (3.9)}$$

where $\theta^{(n)} = [\theta_1^{(n)}, \theta_2^{(n)}, \ldots, \theta_v^{(n)}]$ is the n-th approximation for θ, obtained by solving the equations

$$\theta\left(C + \frac{1}{\rho^{(n-1)}} C_1\right) = Q + \frac{1}{\rho^{(n-1)}} Q_1. \qquad \ldots \text{ (3.10)}$$

As a first approximation for θ we may take its intra-block estimate.

The asymptotic variance of $\hat{\rho}$ obtained from the information matrix is :

$$V(\hat{\rho}) \doteq \frac{2k}{b(k-1)} \rho^2 \qquad \qquad \dots \ (3.11)$$

The right hand side of (3.11) serves as a lower bound for the variance of any unbiased estimator of ρ.

4. QUADRATIC ESTIMATORS σ_0^2 AND σ_1^2

Since the maximum likelihood estimates are somewhat difficult to compute, we may restrict ourselves to quadratic estimators for σ_0^2 and σ_1^2. We notice that the transformed variables z_{0s} ($s = 1, 2, ..., e_0$), $z_{1s}(s = 1, 2, ..., e_1)$ and $z_s(s = 1, 2, ..., q)$ defined by (2.15) each have expectation zero and they are mutually uncorrelated. The variances of z_{0s} and z_{1s}'s are given by (2.8) and (2.9) and the variance of z_s is

$$V(z_s) = \sigma_0^2 + c_s \sigma_1^2 \qquad \qquad \dots \ (4.1)$$

where

$$c_s = a_{0s}^2 / a_{1s}^2 = (rk - \phi_s)/\phi_s. \qquad \qquad \dots \ (4.2)$$

Obviously, we need consider only quadratic forms of the diagonal type

$$Q = b_0 S_0 + b_1 S_1 + \sum_{s=1}^{q} a_s z_s^2 \qquad \qquad \dots \ (4.3)$$

where b_0, b_1 and a_s ($s = 1, 2, ..., q$) are the coefficients to be determined. The expectation of Q is

$$E(Q) = \left(b_0 e_0 + \sum_{s=1}^{q} a_s \right) \sigma_0^2 + \left(b_1 e_1 + \sum_{s=1}^{q} a_s c_s \right) \sigma_1^2. \qquad \dots \ (4.4)$$

The variance of Q, under the assumption that the y_{tu}'s follow a joint normal distribution, is

$$V(Q) = 2 \left[\left(b_0^2 e_0 + \sum_{s=1}^{q} a_s^2 \right) + 2\rho \sum_{s=1}^{q} a_s^2 c_s + \rho^2 \left(b_1^2 e_1 + \sum_{s=1}^{q} a_s^2 c_s^2 \right) \right] \sigma_0^4. \qquad \dots \ (4.5)$$

It is therefore possible to choose b_0, b_1 and a_s ($s = 1, 2, ..., q$) so as to make Q an unbiased estimator of σ_0^2 (or of σ_1^2) with a variance which is minimum for a given value of $\rho = \sigma_1^2/\sigma_0^2$. This gives, for estimating σ_0^2

$$b_0 = (e_1 + B_1)/\Delta, \quad b_1 = -B_0/\Delta \qquad \qquad \dots \ (4.6)$$

and for estimating σ_1^2

$$b_0 = -A_1/\Delta, \qquad b_1 = (e_0 + A_0)/\Delta \qquad \qquad \dots \ (4.7)$$

and in either case

$$a_s = (b_0 + \rho^2 b_1 c_s)/(1 + \rho c_s)^2 \qquad \qquad \dots \ (4.8)$$

where

$$A_0 = \sum_{s=1}^{q} (1+\rho c_s)^{-2}, \quad A_1 = \rho^2 \sum_{s=1}^{q} c_s(1+\rho c_s)^{-2} \qquad \dots \ (4.9)$$

$$B_0 = \sum_{s=1}^{q} c_s(1+\rho c_s)^{-2}, \quad B_1 = \rho^2 \sum_{s=1}^{q} c_s^2(1+\rho c_s)^{-2}$$

and
$$\Delta = (e_0+A_0)(e_1+B_1) - B_0 A_1.$$

The case where ρ is large is of special interest. In such a case the term involving ρ^2 in $V(Q)$ would be dominant, and we may like to minimise this term. The optimum unbiased estimates of σ_0^2 and σ_1^2 in this sense are given by

$$v_0 = S_0/e_0 \qquad \dots \ (4.10)$$

$$v_1 = \frac{S_0}{e_0(b-1)} \sum_{s=1}^{q} \frac{1}{c_s} + \frac{1}{(b-1)} \left[S_1 + \sum_{s=1}^{q} \frac{z_s^2}{c_s} \right] \qquad \dots \ (4.11)$$

respectively. In actual computation we make use of the fact that

$$S_1 + \sum_{s=1}^{q} \frac{z_s^2}{c_s} = \frac{1}{k} \sum_i B_i^2(\theta^*) - \frac{G^2}{bk} \qquad \dots \ (4.12)$$

where $B_i(\theta)$ is defined by (2.17) and θ^* is the intra-block estimate of θ obtained from (1.8). This estimate was suggested by Shah (1962) from intuitive considerations. The variance of v_1 is

$$V(v_1) = \frac{2\sigma_0^4}{(b-1)^2} [(b-1)\rho^2 + 2(\alpha_{-1}-q)\rho + \alpha_{-2} - 2\alpha_{-1} + q + (\alpha_{-1}-q)^2/e_0] \qquad \dots \ (4.13)$$

where
$$\alpha_t = \sum_{s=1}^{q} \left(1 - \frac{\phi_s}{rk} \right)^t \qquad \dots \ (4.14)$$

We may compare this with the customary estimate s_1^2 of σ_1^2 as defined by (1.11). The variance of this estimator is

$$V(s_1^2) = \frac{2k^2\sigma_0^4}{v^2(r-1)^2} \left[(e_1+\alpha_2)\rho^2 + 2(\alpha_1-\alpha_2)\rho \right.$$

$$\left. + 1 - 2\alpha_1 + \alpha_2 + \left(\frac{v}{k} - 1 \right)^2 / e_0 \right]. \qquad \dots \ (4.15)$$

5. Unbiased estimators of ρ

As a convenient unbiased estimator of ρ we may consider a statistic of the form

$$P = \frac{aS_1 + \sum b_s z_s^2}{S_0} + c \qquad \dots \ (5.1)$$

where $a, b_s (s = 1, 2, \dots, q)$ and c are constants to be suitably determined. Since for $e_0 > 2$

$$E(P) = \frac{ae_1\rho + \sum b_s(1+\rho c_s)}{e_0-2} + c,$$

277

to make P an unbiased estimator of ρ, we must have

$$\frac{\Sigma b_s}{e_0-2} +c = 0,$$

and

$$ae_1+\Sigma b_s c_s = e_0-2. \qquad \ldots (5.2)$$

If $e_0 > 4$ the variance of such an unbiased estimator turns out to be

$$V(P) = A_0 + A_1\rho + A_2\rho^2, \qquad \ldots (5.3)$$

where

$$A_0 = 3 \Sigma b_s^2 + 2ae_1 \Sigma b_s - \left(\frac{\Sigma b_s}{e_0-2} \right)^2$$

$$A_1 = \frac{6 \Sigma b_s^2 c_s + 2ae_1 \Sigma b_s c_s}{(e_0-2)(e_0-4)} - \frac{2\Sigma b_s}{e_0-2}$$

$$A_2 = \frac{a^2 e_1(e_1+2) + 3\Sigma b_s^2 c_s^2}{(e_0-2)(e_0-4)} - 1.$$

If we like to minimise A_2 the coefficient of ρ^2 in (5.3), we have to take

$$a = \frac{3(e_0-2)}{3e_1+(e_1+2)q} , \quad b_s = \frac{(e_1+2)a}{3c_s} . \qquad \ldots (5.4)$$

It can be seen that R given by (1.12) is *not* an unbiased estimator of ρ, but a simple correction can be applied to it to make it unbiased. We thus get

$$\left(1-\frac{2}{e_0} \right) R - \frac{2(v-k)}{e_0 v(r-1)} \qquad \ldots (5.5)$$

as an unbiased estimator of ρ.

Similarly, if we start with v_0 and v_1 defined by (4.10) and (4.11) as estimators of σ_0^2 and σ_1^2 respectively, we get, as another unbiased estimator of ρ :

$$\left(1-\frac{2}{e_0} \right) \frac{v_1}{v_0} - \frac{2(v-1)}{e_0(b-1)} \left(\frac{1}{E}-1 \right), \qquad \ldots (5.6)$$

where E is the efficiency-factor of the design (Kempthorne, 1956; Roy, 1958).

$$E = \frac{(v-1)}{rk} \bigg/ \left\{ \frac{v-1-q}{rk} + \sum_{s=1}^{q} \frac{1}{rk-\phi_s} \right\}. \qquad \ldots (5.7)$$

Since with positive probability these unbiased estimators of ρ may turn out to be less than unity, we may use their truncated forms instead, as indicated by (1.13). Let x be any unbiased estimator of ρ and x' its truncated form defined by $x' = 1$ if $x \leqslant 1$, and $x' = x$, otherwise. Then, even though x' is generally a biased estimator for ρ, it can be easily seen that its mean square error can never exceed that for x,

$$E(x'-\rho)^2 \leqslant E(x-\rho)^2.$$

6. SOME PROPERTIES OF COMBINED INTRA AND INTER-BLOCK ESTIMATORS OF TREATMENT EFFECTS

As a corollary to result 2.8, we conclude that when ρ is given, unbiased estimators of treatment effects with minimum variance are obtained from the equations $\tau_s = \bar{t}_s(\rho)$, $s = 1, 2, ..., v-1$, where the right-hand side is given by (2.11). In this section we shall investigate the properties of the estimators of treatment effects obtained by substituting some estimator ρ^* for ρ in the above expression. For typographical simplicity, we shall write

$$\bar{t}_s = \bar{t}_s(\rho), \qquad \bar{t}_s^* = \bar{t}_s(\rho^*).$$

Let

$$w_s = \frac{(\rho^* - \rho)z_s}{1 + \rho^* c_s}, \; s = 1, 2, ..., q. \qquad ... \quad (6.1)$$

We then have the following :

Lemma 6.1 : *If ρ^* satisfies the conditions*

$$E(w_s) = 0, \; V(w_s) < \infty \qquad ... \quad (6.2)$$

for all values of ρ, then $E(\bar{t}_s^*) = \tau_s$

and

$$V(\bar{t}_s^*) = V(\bar{t}_s) + \frac{c_s^2}{a_{1s}^2(1 + \rho c_s)^2} \, V(w_s). \qquad ... \quad (6.3)$$

To prove this, we note that

$$\bar{t}_s^* = \bar{t}_s + \frac{c_s}{a_{1s}(1 + \rho c_s)} \, w_s. \qquad ... \quad (6.4)$$

Also, when ρ is given, $\bar{t}_s$ is the unbiased minimum variance estimator of and by the conditions of the lemma w_s is a zero-function. By Stein's theorem (1950) $\bar{t}_s$ and w_s are uncorrelated. Hence the lemma.

Let P be any statistic of the form (5.1) and let ρ^* be defined as

$$\rho^* = \begin{cases} P & \text{if } P \geqslant 1 \\ 1 & \text{otherwise.} \end{cases} \qquad ... \quad (6.5)$$

It can then be shown that ρ^* so defined satisfies conditions (6.2). That $E(w_s^2)$ is finite can be easily checked. To show that $E(w_s) = 0$, we note that ρ^* is an even function of z_{0s}, $(s = 1, 2, ..., e_0)$, z_{1s} $(s = 1, 2, ..., e_1)$ and z_s $(s = 1, 2, ..., q)$ and consequently, w_s is an odd function of these variables. Since the z's are mutually independent random variables each having a normal distribution with mean zero, the result follows. This is merely an extension of Graybill and Weeks (1959) argument for balanced designs to the case of general incomplete block designs. A similar argument gives

$$E(w_s w_{s'}) = 0$$

for $s \neq s' = 1, 2, ..., q$. Since $\bar{t}_s$ and $\bar{t}_{s'}$ are independent, it follows that $\bar{t}_s^*$ and $\bar{t}_{s'}^*$ are uncorrelated for $s \neq s' = 1, 2, ..., v-1$.

Now, any treatment contrast τ can be expressed as $\tau = \Sigma\ l_s \tau_s$ where $l_s\ (s = 1, 2, ..., v-1)$ are some constants. The minimum variance unbiased estimator of τ when ρ is known is $\bar{l} = \Sigma\ l_s \bar{l}_s$. When ρ is not known, for a combined inter and intra-block estimator of τ, one takes $\bar{l}^* = \Sigma\ l_s \bar{l}_s^*$ by substituting a suitable estimator ρ^* for ρ. If ρ^* satisfies the conditions of Lemma 6.1, we get the following :

Theorem 6.1. *The estimator $\bar{l}^*$ is unbiased for τ, and its variance is given by*

$$V(\bar{l}^*) = V(\bar{l}) + \sum_{s=1}^{f}\ \frac{c_s^2 l_s^2\ V(w_s)}{a_{1s}^2 (1+\rho c_s)^2}$$

the second term being the additional variance due to the sampling fluctuation in ρ^*.

References

Graybill, F. A. and Weeks, D. L. (1959) : Combining inter-block and intra-block information in balanced incomplete blocks. *Ann. Math. Stat.*, **30**, 799-805.

Kempthorne, O. (1956) : The efficiency factor of an incomplete block design. *Ann. Math. Stat.*, **27**, 846-849.

Nair, K. R. (1944) : The recovery of inter-block information in incomplete designs. *Sankhyā*, 6, 383-390.

Rao, C. R. (1947) : General methods of analysis for incomplete block designs. *J. Amer. Stat. Ass.*, **42**, 541-561.

———— On the recovery of inter block information in varietal trials. *Sankhyā*, **17**, 105-114.

Roy, J. (1958) : On the efficiency factor of block designs. *Sankhyā*, **19**, 181-188.

Shah, K. R. (1962) : An alternative estimate of inter-block variance in incomplete block designs. *Sankhyā*, Series A, **24**, 281-286.

Stein, C. (1950) : Unbiased estimates with minimum variance. *Ann. Math. Stat.*, **21**, 406-415.

Weeks, D. L. and Graybill, F. A. (1961) : Minimal sufficient statistics for the balanced incomplete block design under an Eisenhart model II. *Sankhyā*, **23**, 261-268.

Yates, F. (1939) : The recovery of inter-block information in varietal trials arranged in three dimensional lattice. *Ann. Eugen.*, **9**, 136-156.

———— (1940) : The recovery of inter-block information in balanced incomplete block designs. *Ann. Eugen.*, **10**, 317-325.

Paper received : December, 1961.

AN ESTIMATE OF INTER-GROUP VARIANCE
IN ONE AND TWO-WAY DESIGNS

By K. R. SHAH

Indian Statistical Institute

SUMMARY. A procedure for estimation of inter-block variance has been suggested as an alternative to the usual procedure given by Yates and others. A computational procedure to obtain this estimate is also given. The variance of this estimate is compared with that of the usual estimate when ρ, the ratio of the inter-block variance to the intra-block variance is large, say $\rho \geqslant \rho_0$, the estimate considered here has a smaller sampling variance. A table showing these values of ρ_0 is given for all BIB designs listed by Fisher and Yates. Similar procedures for estimating inter-row and inter-column variances are given for designs where the experimental material is symbolically arranged in rows and columns to eliminate heterogeneity in two directions.

1. INTRODUCTION

Yates (1939, 1940) suggested the use of inter-block contrasts for estimating treatment effects. In order to combine these estimates with those obtained from intra-block contrasts, one needs knowledge or at least the estimates of variances of both inter and intra-block contrasts. The variance of a normalised intra-block contrast which we shall call intra-block variance is estimated by the error mean square in the ordinary analysis of variance. The variance of a normalised inter-block contrast which we shall call the inter-block variance was estimated by Yates (1939, 1940), Nair (1944) and Rao (1947, 1956) using the block sum of squares adjusted for treatments.

In the case of two-way designs where the experimental material is symbolically arranged in rows and columns, for the same purpose of utilising inter-row and inter-column contrasts for estimating treatment effects, Roy and Shah (1961) gave procedures for estimating inter-row and inter-column variances. The estimates considered there are analogous to the estimate of inter-block variance given by Yates.

In this paper an alternative estimation procedure is put forth for the estimation of inter-group variance, where the experimental material is grouped in blocks as in one-way designs or in rows and columns as in two-way designs. In the case of two-way designs this procedure turns out to be computationally simpler.

2. PRELIMINARIES

Consider an experiment in which v treatments are tested on bk plots, arranged in b blocks of k plots each, such that each plot receives exactly one treatment, and each treatment is applied atmost once in a block and altogether in r blocks. Let y_{iu} denote the yield of the u-th plot in the i-th block. It is assumed that

$$y_{iu} = \mu + \beta_i + \sum_{j=1}^{u} \theta_j m_{jiu} + e_{iu} \qquad \ldots \quad (2.1)$$

where μ is the general mean, β_i the effect of the i-th block, θ_j the effect of the j-th treatment, $m_{jiu} = 1$ or 0 according as the u-th plot in the i-th block does or does not receive the j-th treatment and ϵ_{iu} is the experimental error, $i = 1, 2, ..., b; j = 1, 2, ..., v; u = 1, 2, ..., k$. The general mean μ and the treatment effects θ_j are regarded as unknown constant parameters, subject to the restriction that $\Sigma\theta_j = 0$. The block effects β_i's and the experimental errors ϵ_{iu}'s are taken to be independent random variables, each with expectation zero and variances given by

$$V(\beta_i) = \sigma_\beta^2, \quad V(\epsilon_{iu}) = \sigma_0^2 \quad \text{for all } (i, u), \qquad \ldots \quad (2.2)$$

we shall write
$$\sigma_1^2 = \sigma_0^2 + k\sigma_\beta^2, \quad \rho = \sigma_1^2/\sigma_0^2. \qquad \ldots \quad (2.3)$$

Let $\sum_u m_{jiu} = n_{ji}$, the number of times the j-th treatment occurs on plots in the i-th block. Thus $n_{ji} = 1$ or 0 and $\sum_i n_{ji} = r, \sum_j n_{ji} = k$. The $v \times b$ matrix $N = ((n_{ji}))$ is called the incidence matrix of the design.

Denote by B_i, the total yield of the i-th block, by T_j, the total yield for the j-th treatment, and by G the total yield of all the plots. Thus

$$B_i = \sum_u y_{iu}, \quad T_j = \sum_i \sum_u m_{jiu} y_{iu}, \quad G = \sum_i \sum_u y_{iu}; \qquad \ldots \quad (2.4)$$

we shall use the row vectors $\boldsymbol{B} = (B_1, B_2, ..., B_b)$ and $\boldsymbol{T} = (T_1, T_2, ..., T_v)$, we further define

$$\boldsymbol{Q} = \boldsymbol{T} - \frac{1}{k}\boldsymbol{BN'}, \quad \boldsymbol{Q}_1 = \frac{1}{k}\boldsymbol{BN'} - \frac{rG}{bk}\boldsymbol{E}_{1v}, \quad \boldsymbol{P} = \boldsymbol{B} - \frac{1}{r}\boldsymbol{TN}$$

$$\boldsymbol{C} = r\boldsymbol{I} - \frac{1}{k}\boldsymbol{NN'}, \quad \boldsymbol{C}_1 = \frac{1}{k}\boldsymbol{NN'} - \frac{r^2}{bk}\boldsymbol{E}_{vv}, \quad \boldsymbol{D} = k\boldsymbol{I} - \frac{1}{r}\boldsymbol{N'N} \qquad \ldots \quad (2.5)$$

where $\boldsymbol{E}_{mm}$ stands for a matrix with m rows and n columns with each element unity. It is easy to verify that

$$E(\boldsymbol{Q}) = \boldsymbol{\theta C}, \quad E(\boldsymbol{Q}_1) = \boldsymbol{\theta C}_1, \quad E(\boldsymbol{P}) = 0,$$

$$V(\boldsymbol{Q}) = \boldsymbol{C}\sigma_0^2, \quad V(\boldsymbol{Q}_1) = \boldsymbol{C}_1\sigma_1^2, \quad V(\boldsymbol{P}) = \boldsymbol{D}\sigma_0^2 + \boldsymbol{D}^2\sigma_\beta^2, \qquad \ldots \quad (2.6)$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_v)$ is the vector of treatment effects parameters.

As is well known (Rao, 1947) the intra-block equations for estimating treatment differences are

$$\boldsymbol{Q} = \boldsymbol{\theta C} \qquad \ldots \quad (2.7)$$

giving us $\hat{\boldsymbol{\theta}} = \boldsymbol{Q C^*}$ as a solution, where we put $\boldsymbol{A^*}$ for the pseudo-inverse (Rao, 1955) of the matrix $\boldsymbol{A}$. It is also shown by Rao (1947) that the combined inter and intra-block equations are

$$\boldsymbol{Q} + \frac{1}{\rho}\boldsymbol{Q}_1 = \boldsymbol{\theta}\left(\boldsymbol{C} + \frac{1}{\rho}\boldsymbol{C}_1\right). \qquad \ldots \quad (2.8)$$

Since ρ is usually unknown, to solve these equations one uses an estimate of ρ obtained by taking the ratio of the estimate of σ_1^2 to that of σ_0^2.[(Yates, 1939, 1940); (Rao, 1947)]. An estimate of σ_0^2 is provided by MS, the error mean square in the ordinary analysis of variance. We now consider the problem of estimating σ_1^2.

3. ESTIMATES OF σ_1^2

Estimate of σ_1^2 using sum of squares due to blocks adjusted for treatments was obtained by Yates (1939, 1940) for special designs and was later adopted by Rao (1947) for general incomplete block designs. This can be put in the form

$$E_1 = \frac{kPD^*P' - (v-k)MS_e}{v(r-1)}. \qquad \dots \quad (3.1)$$

Another estimate is
$$E_2 = \frac{(B-\hat{\theta}N)(I-\frac{1}{b}E_{bb})(B-\hat{\theta}N)' - \alpha MS_e}{k(b-1)} \qquad \dots \quad (3.2)$$

where α is the trace of the matrix C^*NN', which can be shown to be equal to $k(v-1)$ $(1-E)/E$ where E is the efficiency factor of the design (Kempthorne, 1956; Roy, 1958).

Under the additional assumption that the random variables β_i's and ϵ_{iu}'s are jointly normally distributed we shall derive the variances of E_1 and E_2. Now, we state the following lemma used in deriving $V(E_1)$ and $V(E_2)$.

Lemma : If $x = (x_1, x_2, ..., x_n)$ has multivariate normal distribution with mean 0 and dispersion matrix Λ and if Δ is a matrix such that there exists an orthogonal matrix P satisfying

$$P\Lambda P' = \operatorname{diag}(\lambda_1, \lambda_2, ..., \lambda_n), \quad P\Delta P' = \operatorname{diag}(\delta_1, \delta_2, ..., \delta_n)$$

then
$$E(x\Delta x') = \sum_i \delta_i \lambda_i \qquad \dots \quad (3.3)$$

$$V(x\Delta x') = 2\sum_i \delta_i^2 \lambda_i.$$

The proof is immediate if we transform from x to $Z = xP'$.

First, we observe that MS_e is distributed independently of P, B and $\hat{\theta}$. We also note that B and $\hat{\theta}$ are independently distributed. Hence $V(B-\theta\hat{N}) = kI\sigma_1^2 + N'C^*N\sigma_0^2$.

Using the Lemma we get
$$V(E_1) = \frac{2}{v^2(r-1)^2}\left[k^2 \sum_{i=1}^{b-1}(x_i\sigma_\beta^2 + \sigma_0^2)^2 + \frac{1}{e}(v-k)^2\sigma_0^4 \right] \qquad \dots \quad (3.4)$$

and
$$V(E_2) = \frac{2}{k^2(b-1)^2}\left[k^4 \sum_{i=1}^{b-1}(\sigma_\beta^2 + \sigma_0^2/x_i)^2 + \frac{1}{e}(\alpha^2\sigma_0^4) \right]$$

where $x_1, x_2, ..., x_{b-1}$ are the non-zero latent roots of D (when all treatment contrasts are estimable, i.e., when the design is connected, the matrix D has exactly one latent

root zero. Here, we shall consider only connected designs) and e stands for the number of error degrees of freedom in the intra-block analysis.

Since $\bar{x} = \Sigma x_i /(b-1) = v(r-1)/(b-1)$, it is easy to see that

$$V(E_1) - V(E_2) = \frac{2k^2}{(b-1)^2\bar{x}^2} \left[\sum_{i=1} \left\{ (x_i\sigma_\beta^2 + \sigma_0^2)^2 - \bar{x}^2(\sigma_\beta^2 + \sigma_0^2/x_i)^2 \right\} \right.$$

$$\left. + \left\{ \frac{(v-k)^2}{k^2} - \frac{\alpha^2\bar{x}^2}{k^4} \right\} \frac{\sigma_0^4}{e} \right]. \quad \ldots \quad (3.5)$$

On simplification this gives

$$V(E_1) - V(E_2) = \frac{2k^2}{(b-1)^2\bar{x}^2} \left[\sigma_\beta^4 \left\{ \sum_i x_i^2 - (b-1)\bar{x}^2 \right\} + 2\sigma_\beta^2\sigma_0^2 \left\{ \Sigma x_i - \bar{x}^2 \Sigma \frac{1}{x} \right\} \right.$$

$$\left. + \sigma_0^4 \left\{ b-1 - \bar{x}^2 \Sigma \frac{1}{x_i^2} + \frac{(v-k)^2}{k^2e} - \frac{\alpha^2\bar{x}^2}{k^4e} \right\} \right]. \quad \ldots \quad (3.6)$$

In this expression σ_β^4 has positive coefficient while σ_0^4 and $\sigma_0^2\sigma_\beta^2$ have negative coefficients. Hence $V(E_1) - V(E_2)$ is positive for somewhat large values of $\sigma_\beta^2/\sigma_0^2$ or equivalently for somewhat large values of $\rho = \sigma_1^2/\sigma_0^2$. By ρ_0, we denote the value of ρ such that for $\rho > \rho_0$, $V(E_1) - V(E_2)$ is positive i.e., the usual estimate has larger variance than the new estimate. The values of ρ_0 are given below in the table for all BIB designs listed by Fisher and Yates (1957). For each of these designs ρ_0 happens to lie between 4 and 5.

VALUES OF ρ_0 FOR ALL BIBD BY FISHER AND YATES
(Other than Symmetrical Designs)

k	r	b	v	ρ_0	k	r	b	v	ρ_0
3	6	10	5	4.3429	5	10	18	9	4.1795
3	5	10	6	4.3828	5	9	18	10	4.1875
3	4	12	9	4.4179	5	7	21	15	4.2225
3	6	26	13	4.6270	5	6	30	25	4.1900
3	9	30	10	4.6801	5	10	82	41	4.3435
3	7	35	15	4.7080	6	8	12	9	4.0822
3	9	57	19	4.8091	6	9	15	10	4.1062
3	10	70	21	4.8463	6	9	24	16	4.1584
					6	8	28	21	4.1560
4	7	14	8	4.2520	6	9	69	46	4.2142
4	10	15	6	4.2044	6	10	85	51	4.2406
4	6	15	10	4.2640	7	10	30	21	4.1304
4	8	18	9	4.3016	7	9	36	28	4.1290
4	5	20	16	4.2584	7	8	56	49	4.1157
4	8	50	25	4.4464	8	10	45	36	4.1040
4	9	63	28	4.4832	8	9	72	64	4.0976
					9	10	90	81	4.0672

It may be noted that the two estimates are identical in the case of $x_1 = x_2 = \cdots = x_{b-1}$ i.e. when the design is a dual of a BIB design. Hence in particular for symmetrical BIB designs the two estimates are identical.

Other properties of E_2 will be discussed in a subsequent communication. We derive here a simple computational procedure to obtain E_2 after the intra-block analysis is performed. It is easy to see that

$$(B - \hat{\theta}N)\left(I - \frac{1}{b}E_{bb}\right)(B - \theta\hat{N})' = kSS_B^* - BN'\hat{\theta}' + \hat{\theta}NN'\hat{\theta}' \qquad \ldots (3.7)$$

where SS_B^* denotes unadjusted block sum of squares. On simplification this reduces to $k[SS_B^* - (T - r\theta)\hat{\theta}']$. Once, the intra-block estimates $\hat{\theta}$ are obtained, this is easily computed.

4. Designs with two-way elimination of heterogeneity

A similar problem for designs with two-way elimination of heterogeneity would necessitate estimates of three variances. Here, if the experimental material is arranged into a two-way array of rows and columns one has to estimate inter-row variance, inter-column variance in addition to the usual error variance which one may call interaction variance. As usual, the estimate of interaction variance is provided by MS, the error mean square in the ordinary analysis of variance. A method of obtaining estimates of inter-row and inter-column variances was given by Roy and Shah (1961). The estimates given there are analogous to the estimate of inter-block variance given by E_1 and to compute them one has to carry out two additional one-way analysis, one with rows and treatments ignoring columns and the other with columns and treatments ignoring rows. In this case, it would be simpler to compute the estimates of inter-row and inter-column variances corresponding to E_2 in the one-way case.

If σ_R^2 denotes the variance of a normalised inter-row contrast an estimate of σ_R^2 is given by

$$\hat{\sigma}_R^2 = \frac{(R - tM)\left(I - \frac{1}{m}E_{mm}\right)(R - tM)' - tr(K^*MM')MS_e}{n(m-1)} \qquad \ldots (4.1)$$

where all the quantities defined in the R.H.S. are as defined by Roy and Shah (1961). Using the same notations, the estimate of σ_c^2, the variance of a normalised inter-column contrast is given by

$$\hat{\sigma}_C^2 = \frac{(C - tN)\left(I - \frac{1}{n}E_{nn}\right)(C - tN)' - tr(K^*NN')MS_e}{m(n-1)} . \qquad \ldots (4.2)$$

After the interaction analysis is performed $\hat{\sigma}_R^2$ and $\hat{\sigma}_c^2$ can be easily calculated.

References

Fisher, R. A. and Yates, F. (1957): *Statistical Tables for Biological, Agricultural and Medical Research*, 5th ed. Edinburgh, Oliver and Boyd.

Kempthorne, O. (1956): The efficiency factor of an incomplete block design. *Ann. Math. Stat.*, 27, 846-849.

Nair, K. R. (1944): The recovery of inter-block information in incomplete block designs. *Sankhyā*, 6, 383-390.

Rao, C. R. (1947): General methods of analysis for incomplete block designs. *J. Amer. Stat. Ass.*, 42, 541-561.

———— (1955): Analysis of dispersion for multiply classified data with unequal numbers in cells. *Sankhyā*, 15, 253-280.

———— (1956): On the recovery of inter-block information in varietal trials. *Sankhyā*, 17, 105-114.

Roy, J. (1958): On the efficiency factor of block designs. *Sankhyā*, 19, 181-188.

Roy, J. and Shah, K. R. (1961): Analysis of two-way designs. *Sankhyā*, 23, 129-144.

Yates, F. (1939): The recovery of inter-block information in varietal trials arranged in three dimensional lattice. *Ann. Eug.*, 9, 136-156.

———— (1940): The recovery of inter-block information in balanced incomplete block designs. *Ann. Eug.* 10, 317-325.

Paper received : January, 1962.

ON SIMPLE RANDOM SAMPLING WITH REPLACEMENT

By P. K. PATHAK

Indian Statistical Institute

SUMMARY. In simple random sampling with replacement, Basu (1958), and Des Raj and Khamis (1958), showed that for estimating the population mean, the average of distinct units is more efficient than the overall sample mean. In this paper, a detailed treatment of the above problem is given, and the exact expression for the variance of above estimator is derived. The relative efficiency of the above estimator with other estimators is also considered. An improved estimator of the population variance is obtained. Finally, a comparison between the two simple random sampling schemes (with and without replacement) is made.

1. INTRODUCTION

We index the N population units as $1, 2, ..., N$, and let Y_j be some real-valued characteristic (in which we are interested) of the j-th population unit.* Here we consider the problem of estimating the population mean

$$\bar{Y} = N^{-1} \Sigma\ Y_j$$

and the population variance $\quad \sigma^2 = N^{-1} \Sigma\ (Y_j - \bar{Y})^2.$

For simplicity we refer population units by capital letters and sample units by small letters, e.g., u_i and y_i will denote the unit index and the variate value respectively associated with the i-th sample unit.

2. ESTIMATION OF $\bar{Y}$

In simple random sampling (with replacement), Basu (1958) considered two estimators of the population mean

(i) $\bar{y} = 1/n\ \Sigma\ y_i \equiv$ average of n sample units;

(ii) $\bar{y}_\nu = 1/\nu \Sigma y_{(i)} \equiv$ average of ν distinct units observed in the sample.

If we record the sample of observation as

$$S = (x_1, x_2, ..., x_n),$$

where $x_i = (y_i, u_i)$; and if ν be the number of distinct units observed in the sample, Basu (1958) showed that the 'order-statistic' (sample units arranged in ascending order of their unit-indices)

$$T = [x_{(1)}, x_{(2)}, ..., x_{(\nu)}]$$

(where $x_{(i)} = (y_{(i)}, u_{(i)})$, and $y_{(i)}$ is the variate value of the sample unit with unit index $u_{(i)}$) forms a sufficient statistic, and therefore, for any convex (downwards) loss function

$$E(\bar{y}|T) = E(y_1|T) = \bar{y}_\nu, \qquad \qquad ...\ (2.1)$$

has uniformly smaller risk than $\bar{y}$.

An exact expression for variance of $\bar{y}_\nu$ is given below.

Variance of $\bar{y}_\nu$. We have

$$V(\bar{y}_\nu) = E\{V(\bar{y}_\nu|\nu)\} = E\left(\frac{1}{\nu} - \frac{1}{N}\right) S^2 \qquad ...\ (2.2)$$

where $\quad S^2 = [N/(N-1)]\sigma^2.$

---

* j runs from 1 to N; i from 1 to n; and (i) from (1) to (ν).

11

Since
$$E\left(\frac{1}{\nu}\right) = \frac{1^{n-1}+2^{n-1}+\ldots+N^{n-1}}{N^n}, \quad \text{(Pathak, 1961)}$$

$$V(\bar{y}_\nu) = \frac{1^{n-1}+2^{n-1}+\ldots+(N-1)^{n-1}}{N^n}\, S^2. \quad \ldots \; (2.3)$$

For large samples, it is rather cumbersome to compute $V(\bar{y}_\nu)$. An approximate expression for $V(\bar{y}_\nu)$ valid for terms up to order N^{-2} is given by

$$V(\bar{y}_\nu) \doteqdot \left[\frac{1}{n} - \frac{1}{2N} + \frac{(n-1)}{12N^2}\right] S^2. \quad \ldots \; (2.4)$$

3. Admissibility properties of certain estimators of $\bar{Y}$

Let Γ denote a certain class of estimators of $\bar{Y}$. For a given loss function, let $R(t)$ represent the risk (or expected loss) associated with the estimator t of $\bar{Y}$.

Of the two estimators t_1 and t_2 of $\bar{Y}$, t_1 will be said to be uniformly better than t_2 if, for a given loss function,

$$R(t_1) \leqslant R(t_2) \quad \ldots \; (3.1)$$

holds for all possible values of $(Y_1, Y_2, \ldots, Y_N)$ with strict sign of inequality holding for at least one $(Y_1, Y_2 \ldots, Y_N)$.

An estimator t belonging to Γ is said to be admissible in Γ if there exists no estimator in Γ which is better than t.

Now we consider the problem of finding admissible estimators of $\bar{Y}$. As the 'order-statistic' T is sufficient, we have to restrict ourselves to functions of T only. Moreover, the distribution of T is not complete, therefore, many different estimators of $\bar{Y}$ can be suggested. For simplicity, we shall consider the following class of unbiased linear estimators of $\bar{Y}$.

$$\bar{y}_s = f_1(\nu)\, \bar{y}_\nu + f_2(\nu). \quad \ldots \; (3.2)$$

In view of the fact that $\quad E[\bar{y}_s \,|\, \nu] = f_1(\nu)\bar{Y} + f_2(\nu),$

obviously, necessary and sufficient conditions for $\bar{y}_s$ to be an unbiased estimator of $\bar{Y}$, are

$$E[f_1(\nu)] = 1 \quad \text{and} \quad E[f_2(\nu)] = 0. \quad \ldots \; (3.3)$$

Consider now the class Γ of estimators $\bar{y}_s$ which satisfy the conditions of (3.3).

Now
$$V(\bar{y}_s) = E\left[f_1^2(\nu)\left(\frac{1}{\nu} - \frac{1}{N}\right)S^2\right] + V[f_1(\nu)\bar{Y} + f_2(\nu)]. \quad \ldots \; (3.4)$$

In order to choose a good estimator from Γ, we are to minimise (3.4) by proper choices of $f_1(\nu)$ and $f_2(\nu)$. The first expression on the right hand side of (3.4) is independent of $f_2(\nu)$; so, for a proper choice of $f_2(\nu)$, we are to minimise

$$V[f_1(\nu)\bar{Y} + f_2(\nu)]$$

which is minimum if $\bar{Y}f_1(\nu) + f_2(\nu)$ is constant for all values of ν, i.e.,

$$\bar{Y}f_1(\nu) + f_2(\nu) = E[f_1(\nu)\bar{Y} + f_2(\nu)] = \bar{Y},$$

or
$$f_2(\nu) = \bar{Y}[1 - f_1(\nu)]. \quad \ldots \; (3.5)$$

Since the above solution of $f_2(\nu)$ contains the unknown $\bar{Y}$, the exact value of $f_2(\nu)$ is not known unless $f_1(\nu) = 1$. Thus, if we choose $f_1(\nu) = 1$, the best estimator of $\bar{Y}$ would be $\bar{y}_\nu$. However, in practical situations, when some *a priori* knowledge about $\bar{Y}$ is available, it seems appropriate to approximate $f_2(\nu)$ by

$$f_2(\nu) = \bar{X}[1-f_1(\nu)], \qquad \dots (3.6)$$

where $\bar{X}$ is some *a priori* estimate of $\bar{Y}$. For example, $\bar{X}$ may be taken as the estimate of the population mean of the same variate obtained from some previous survey etc. On the other hand, if no such information about $\bar{Y}$ is available, it would be safe to take $f_2(\nu) = 0$. To choose the optimum value of $f_1(\nu)$, we have to minimise

$$E f_1^2(\nu)\left(\frac{1}{\nu}-\frac{1}{N}\right)$$

subject to the condition that $E[f_1(\nu)] = 1$.

By Schwartz inequality we have

$$E\left[f_1^2(\nu)\left(\frac{1}{\nu}-\frac{1}{N}\right)\right] \cdot E\left[\left(\frac{1}{\nu}-\frac{1}{N}\right)^{-1}\right] \geqslant 1. \qquad \dots (3.7)$$

The equality holds if and only if

$$f_1(\nu) = \left(\frac{1}{\nu}-\frac{1}{N}\right)^{-1}\Big/\left[E\left(\frac{1}{\nu}-\frac{1}{N}\right)^{-1}\right] = [N\nu/(N-\nu)]/E[N\nu/(N-\nu)]. \qquad \dots (3.8)$$

Thus, when some *a priori* estimate $\bar{X}$ of $\bar{Y}$ is available, the optimum estimate of $\bar{Y}$ is given by

$$\bar{y}_{\nu(1)} = \frac{[N\nu/(N-\nu)]}{E[N\nu/(N-\nu)]}\,\bar{y}_\nu + \bar{X}\left[1-\frac{[N\nu/(N-\nu)]}{E[N\nu/(N-\nu)]}\right]. \qquad \dots (3.9)$$

When no such information about $\bar{Y}$ is available we may use the following estimator

$$\bar{y}_{\nu(2)} = \frac{[N\nu/(N-\nu)]}{E[N\nu/(N-\nu)]}\,\bar{y}_\nu. \qquad \dots (3.10)$$

The two estimators are admissible in Γ in the sense that they minimise the first component of (3.4). Any estimator $\bar{y}_s$ different from either of them cannot be uniformly better than $\bar{y}_{\nu(1)}$ or $\bar{y}_{\nu(2)}$ because

$$V(\bar{y}_{\nu(1)}) < V(\bar{y}_s) \quad \text{for all populations where } \bar{Y} = \bar{X};$$
$$V(\bar{y}_{\nu(2)}) < V(\bar{y}_s) \quad \text{for all populations where } \bar{Y} = 0.$$

Expression for $E[N\nu/(N-\nu]$: Proceeding on similar lines given by the author (Pathak, 1961), it can be shown that

$$E[N\nu/(N-\nu)] = N^2 \sum_{m=1}^{n} \frac{p_{12\dots m}}{(N-m)} \qquad \dots (3.11)$$

where
$$p_{12\ldots m} = \begin{cases} 1- \binom{m}{1}(1-1/N)^n + \ldots (-)^m \binom{m}{m}(1-m/N)^n & \text{for } m \leqslant n; \\ 0 & \text{otherwise.} \end{cases}$$

Thus, we see that it may be quite cumbersome to compute the estimators (3.9) and (3.10) in case of large samples owing to the difficulty of computing $E[N\nu/(N-\nu)]$. If, however, the sampling fraction n/N can be ignored, the estimators reduce to

$$\bar{y}^*_{\nu(1)} = \frac{\nu}{E(\nu)}\,\bar{y}_\nu + \bar{X}\left[1 - \frac{\nu}{E(\nu)}\right]; \qquad \ldots \ (3.12)$$

$$\bar{y}^*_{\nu(2)} = \frac{\nu}{E(\nu)}\,\bar{y}_\nu. \qquad \ldots \ (3.13)$$

It is easy to see that (3.13) is the well-known Horvitz-Thompson (1952) estimator in case of equal probability sampling. An interesting comparison between $V(\bar{y}^*_{\nu(2)})$ and $V(\bar{y}_\nu)$ is made below.

4. Comparison between $(V\bar{y}_\nu)$ and $V(\bar{y}^*_{\nu(2)})$

We have shown that

$$V(\bar{y}_\nu) = \frac{1^{n-1}+2^{n-1}+\ldots+(N-1)^{n-1}}{N^n}\,S^2;$$

and

$$V(\bar{y}^*_{\nu(2)}) = E\left[V\left\{\frac{\nu}{E(\nu)}\bar{y}_\nu | \nu\right\}\right] + V\left[E\left\{\frac{\nu}{E(\nu)}\bar{y}_\nu | \nu\right\}\right]$$

$$= E\left[\frac{\nu^2}{E^2(\nu)}\left(\frac{1}{\nu}-\frac{1}{N}\right)S^2\right] + \frac{\bar{Y}^2}{E^2(\nu)}V(\nu). \qquad \ldots \ (4.1)$$

It can be shown that

$$E(\nu) = N\left[1-\left(1-\frac{1}{N}\right)^n\right];$$

$$E(\nu^2) = N\left[1-\left(1-\frac{1}{N}\right)^n\right] + N(N-1)\left[1-2\left(1-\frac{1}{N}\right)^n+\left(1-\frac{2}{N}\right)^n\right];$$

and
$$V(\nu) = N\left(1-\frac{1}{N}\right)^n - N^2\left(1-\frac{1}{N}\right)^{2n} + N(N-1)\left(1-\frac{2}{N}\right)^n$$

$$\therefore \quad V(\bar{y}^*_{\nu(2)}) = \frac{S^2}{N^2\left[1-\left(1-\frac{1}{N}\right)^n\right]^2}\left[N\left\{1-\left(1-\frac{1}{N}\right)^n\right\}-\left\{1-\left(1-\frac{1}{N}\right)^n\right\}\right.$$

$$-(N-1)\left\{1-2\left(1-\frac{1}{N}\right)^n+\left(1-\frac{2}{N}\right)^n\right\}\right] + \frac{\bar{Y}^2}{N^2\left[1-\left(1-\frac{1}{N}\right)^n\right]^2}$$

$$\times\left[N\left(1-\frac{1}{N}\right)^n - N^2\left(1-\frac{1}{N}\right)^{2n} + N(N-1)\left(1-\frac{2}{N}\right)^n\right] \ldots \ (4.2)$$

Now

$$V(\bar{y}_\nu) - V(\bar{y}_{\nu(2)}^*) = S^2 \left[\frac{1^{n-1} + 2^{n-1} + \dots + (N-1)^{n-1}}{N^n} - \frac{(N-1)\left[\left(1-\frac{1}{N}\right)^n - \left(1-\frac{2}{N}\right)^n\right]}{N^2 \left[1-\left(1-\frac{1}{N}\right)^n\right]^2} \right]$$

$$-\frac{\bar{Y}^2}{N} \cdot \frac{\left[\left(1-\frac{1}{N}\right)^n - N\left(1-\frac{1}{N}\right)^{2n} + (N-1)\left(1-\frac{2}{N}\right)^n\right]}{\left[1-\left(1-\frac{1}{N}\right)^n\right]^2}$$

$$= C_1 S^2 - C_2 \bar{Y}^2. \quad \text{(say)} \qquad \dots \quad (4.3)$$

Thus $\bar{y}_\nu$ is better than $\bar{y}_{\nu(2)}^*$ if

$$\frac{S^2}{\bar{Y}^2} < \frac{C_2}{C_1}$$

and worse if

$$\frac{S^2}{\bar{Y}^2} > \frac{C_2}{C_1}.$$

Approximate values of C_1 and C_2 for large populations correct up to terms of order N^{-2}, are given by

$$C_1 = \frac{1}{2nN} + \frac{5(n-1)}{12nN^2} ;$$

$$C_2 = \frac{(n-1)}{2nN} - \frac{(n-1)(n-2)}{3nN^2}. \qquad \dots \quad (4.4)$$

The above comparison shows that if the square of the population coefficient of variation exceeds $(n-1)$, then $\bar{y}_{\nu(2)}^*$ has smaller variance than $\bar{y}_\nu$. Moreover, if we have some *a priori* knowledge of $\bar{Y}$, it would be more pertinent to compare $\bar{y}_\nu$ and $\bar{y}_{\nu(1)}^*$. It can be seen on similar lines that $\bar{y}_\nu$ is better than $\bar{y}_{\nu(2)}^*$ if

$$\frac{S^2}{(\bar{Y}-\bar{X})^2} < \frac{C_2}{C_1},$$

and worse otherwise. This result shows that if $\bar{X}$ provides a close approximation to $\bar{Y}$, it is always better to use $\bar{y}_{\nu(1)}^*$ rather than $\bar{y}_\nu$.

We now state the following admissibility property of $\bar{y}_\nu$.

Theorem 1 : *If squared error be the loss function, $\bar{y}_\nu$ is admissible among all functions of $\bar{y}_\nu$ and* ν.

Proof : Let $\qquad t = \bar{y}_\nu + f(\bar{y}_\nu, \nu)$

be a function of $\bar{y}_\nu$ and ν. Suppose that t is uniformly better than $\bar{y}_\nu$. Now by hypothesis,

$$R(t) = E(\bar{y}_\nu - \bar{Y})^2 + E[f(\bar{y}_\nu, \nu)]^2 + 2E[(\bar{y}_\nu - \bar{Y})f(\bar{y}_\nu, \nu)] \leqslant E(\bar{y}_\nu - \bar{Y})^2 \quad \dots \quad (4.5)$$

holds for all $Y_1, Y_2, \dots, Y_N$. Take in particular $Y_1 = Y_2 = \dots = Y_N = C$ (say). Then the above relation implies that

$$f(C, \nu) = 0. \qquad \dots \quad (4.6)$$

Since the choice of C is arbitrary, it follows that $f(\bar{y}_\nu, \nu)$ is identically zero, which proves the above theorem.

5. Estimation of Variance

We now turn to the problem of estimating the population variance from a simple random sample (with replacement). The usual estimator of the population variance

$$\sigma^2 = N^{-1} \Sigma (Y_j - \bar{Y})^2$$

is given by the sample variance

$$s^2 = \frac{1}{(n-1)} \Sigma (y_i - \bar{y})^2 = \frac{1}{2n(n-1)} \sum_{i \neq i'} (y_i - y_{i'})^2. \qquad \dots \quad (5.1)$$

In this section, we derive an estimator uniformly better than s^2.

Theorem 2 : *For any convex (downwards) loss function, an estimator uniformly better than s^2 is given by*

$$s_\nu^2 = \left[\frac{C_\nu(n) - C_\nu(n-1)}{C_\nu(n)} \right] s_d^2, \qquad \dots \quad (5.2)$$

where

$$C_\nu(n) = \nu^n - \binom{\nu}{1}(\nu-1)^n + \dots (-)^{\nu-1} \binom{\nu}{\nu-1} 1^n,$$

and

$$s_d^2 = \begin{cases} \dfrac{1}{(\nu-1)} \Sigma (y_{(i)} - \bar{y}_\nu)^2 & \text{if } \nu > 1 \text{ ;} \\[2mm] 0 & \text{otherwise.} \end{cases}$$

Proof : Since the 'order-statistic', T, is sufficient, by Rao-Blackwell theorem, an estimator uniformly better than s^2 is given by

$$E[s^2 | T] = E\left[\frac{1}{2n(n-1)} \sum_{i \neq i'} (y_i - y_{i'})^2 | T \right] = E[\tfrac{1}{2}(y_1 - y_2)^2 | T]. \qquad \dots \quad (5.3)$$

When $\nu = 1$, (5.3) is obviously zero. To derive (5.3) when $\nu > 1$, we observe that

$$P[x_1 = x_{(i)}, x_2 = x_{(i')} | T] = \frac{\dfrac{1}{N^2} \Sigma''' \dfrac{(n-2)!}{\alpha_{(1)}! \dots \alpha_{(\nu)}!} \left(\dfrac{1}{N}\right)^{\alpha_{(1)}} \dots \left(\dfrac{1}{N}\right)^{\alpha_{(\nu)}}}{\Sigma' \dfrac{n!}{\alpha_{(1)}! \dots \alpha_{(\nu)}!} \left(\dfrac{1}{N}\right)^{\alpha_{(1)}} \dots \left(\dfrac{1}{N}\right)^{\alpha_{(\nu)}}} \qquad \dots \quad (5.4)$$

$$(i \neq i' = 1, 2, \dots, \nu)$$

where Σ' means summation over all integral α's such that

$$\alpha_{(1)} + \alpha_{(2)} + \dots + \alpha_{(\nu)} = n \quad \text{and} \quad \alpha_{(i)} > 0 \quad \text{for} \quad i = 1, 2, \dots, \nu;$$

and Σ''' means summation over all integral α's such that

$$\alpha_{(1)} + \alpha_{(2)} + \dots + \alpha_{(\nu)} = n-2, \ \alpha_{(i)} \geqslant 0, \ \alpha_{(i)}' \geqslant 0 \text{ and } \alpha_{(k)} > 0 \text{ ; for } k \neq i \neq i' = 1, 2, \dots, \nu.$$

It follows from Lemma 1 given by the author (Pathak, 1961) that

$$\Sigma' \frac{n!}{\alpha_{(1)}! \dots \alpha_{(\nu)}!} = C_\nu(n);$$

$$\Sigma''' \frac{(n-2)!}{\alpha_{(1)}! \dots \alpha_{(\nu)}!} = C_\nu(n-2) + 2C_{\nu-1}(n-2) + C_{\nu-2}(n-2)$$

$$= \frac{C_\nu(n) - C_\nu(n-1)}{\nu(\nu-1)} \qquad \dots \quad (5.5)$$

$$\therefore P[x_1 = x_{(i)}, \ x_2 = x_{(i')} | T] = \frac{C_\nu(n) - C_\nu(n-1)}{\nu(\nu-1) \ C_\nu(n)} \qquad \dots \quad (5.6)$$

$$(i \neq i' = 1, 2, \dots, \nu).$$

Thus, if $\nu > 1$

$$E\left[\frac{(y_1 - y_2)^2}{2} \Big| T \right] = \Sigma \frac{(y_{(i)} - y_{(i')})^2}{2} \ P[x_1 = x_{(i)}, x_2 = x_{(i')} | T]$$

$$= \frac{C_\nu(n) - C_\nu(n-1)}{C_\nu(n)} \left[\frac{1}{2\nu(\nu-1)} \ \Sigma \ (y_{(i)} - y_{(i')})^2 \right]$$

$$= \frac{C_\nu(n) - C_\nu(n-1)}{C_\nu(n)} \cdot \frac{1}{(\nu-1)} \ \Sigma \ (y_{(i)} - \bar{y}_\nu)^2. \qquad \dots \quad (5.7)$$

Therefore, for any ν, $\quad E(s^2|T) = E\left[\frac{(y_1 - y_2)^2}{2} \Big| T \right] = \frac{C_\nu(n) - C_\nu(n-1)}{C_\nu(n)} \ s_d^2, \qquad \dots \quad (5.8)$

where s_d^2 has been defined earlier.

In practice the estimator s_ν^2 requires the knowledge of the ratio $\frac{C_m(n-1)}{C_m(n)}$.
Table 3 gives values of $\frac{C_m(n-1)}{C_m(n)}$ correct to seven places of decimals for $1 \leqslant m \leqslant n \leqslant 50$;
and were computed from values of $\frac{C_m(n)}{m!}$ tabulated by Gupta (1950).

The following results are direct consequences of Theorem 2.

If there are two characters Y and Z, the covariance between Y and Z is defined by

$$\sigma_{(yz)} = \frac{1}{N} \Sigma(Y_j - \bar{Y})(Z_j - \bar{Z}). \qquad \dots \quad (5.9)$$

The usual estimator of σ_{yz} is given by

$$s_{(yz)} = \frac{1}{(n-1)} \Sigma(y_i - \bar{y})(z_i - \bar{z}). \qquad \dots \quad (5.10)$$

Corollary 1 :   *It follows from Theorem 2 that an estimator better than $s_{(yz)}$ is given by*

$$s_{\nu(yz)} = \begin{cases} \dfrac{C_\nu(n) - C_\nu(n-1)}{C_\nu(n)} \, s_{d(yz)} & if \quad \nu > 1 \\ 0 & otherwise, \end{cases} \qquad ... \quad (5.11)$$

where $s_{d(yz)}$ is the sample covariance based on the distinct units observed in the sample.

The above theorem can be used to derive an unbiased ratio estimator which is better than Hartley-Ross unbiased ratio estimator (1954). In the sampling scheme under consideration, Hartley-Ross estimator is given by

$$\bar{y}_R = \bar{r}\bar{Z} - \frac{n}{(n-1)}(\bar{y} - \bar{r}\bar{z}) = \bar{r}\bar{Z} - \frac{1}{(n-1)} \Sigma \left(\frac{y_i}{z_i} - \bar{r} \right)(z_i - \bar{z}),$$

where $\bar{Z} = N^{-1} \Sigma Z_j$, $\bar{r} = 1/n \, \Sigma y_i/z_i$ and z_i is the value of the Z-characteristic, an auxiliary characteristic related to Y-characteristic, of the i-th sample unit.

Corollary 2 :   *An estimator better than $\bar{y}_R$ is given by*

$$E[\bar{y}_R | T] = \begin{cases} \bar{r}_\nu \bar{Z} + \dfrac{C_\nu(n) - C_\nu(n-1)}{C_\nu(n)} \cdot \dfrac{\nu}{(\nu-1)}(\bar{y}_\nu - \bar{r}_\nu \bar{z}_\nu) & if \quad \nu > 1 \\ \bar{r}_\nu \bar{Z} & otherwise, \end{cases} \qquad ... \quad (5.12)$$

where $\qquad \bar{r}_\nu = \dfrac{1}{\nu} \Sigma \dfrac{y_{(i)}}{z_{(i)}}, \quad \bar{z}_\nu = \dfrac{1}{\nu} \Sigma z_{(i)}.$

Murthy (1961) has extended the idea of ratio estimators to product estimators. Similar to the well-known ratio estimator $\dfrac{\bar{y}\bar{Z}}{\bar{z}}$ he has considered the product estimator

$$\frac{\bar{y}\bar{z}}{\bar{Z}} \qquad ... \quad (5.13)$$

for estimating $\bar{Y}$.

Corollary 3 :   *It can be verified that an estimator better than $\dfrac{\bar{y}\bar{z}}{\bar{Z}}$ is given by*

$$E\left[\frac{\bar{y}\bar{z}}{\bar{Z}} \Big| T \right] = \frac{1}{\bar{Z}} \left[\frac{\Sigma y_{(i)} z_{(i)}}{\nu} - \frac{(n-1)}{n} \cdot s_{\nu(yz)} \right] \qquad ... \quad (5.14)$$

where $s_{\nu(yz)}$ is given by (5.11).

Finally, the almost unbiased product estimator of Murthy (1961), is given which makes $\dfrac{\bar{y}\bar{z}}{\bar{Z}}$ almost unbiased. This estimator is given by

$$P_0 = \frac{n}{(n-1)} \cdot \frac{\bar{y}\bar{z}}{\bar{Z}} - \frac{1}{(n-1)} \frac{\Sigma y_i z_i}{n\bar{Z}}. \qquad ... \quad (5.15)$$

Corollary 4 :   *An estimator better than P_0 is given by*

$$E[P_0 | T] = \frac{\Sigma x_{(i)} \, y_{(i)}}{\nu \bar{Z}} - \frac{s_{\nu(yz)}}{\bar{Z}}. \qquad ... \quad (5.16)$$

6. Some estimators of $V(\bar{y}_\nu)$

Some unbiased estimators of $V(\bar{y}_\nu)$ are given by

(I) $\qquad v_1(\bar{y}_\nu) = \left[\dfrac{1^{n-1}+2^{n-1} \ldots +(N-1)^{n-1}}{N^n} \right] \dfrac{N}{(N-1)} s^2;$

(II) $\qquad v_2(\bar{y}_\nu) = \left[\dfrac{1^{n-1}+2^{n-1}+\ldots+(N-1)^{n-1}}{N^n} \right] \dfrac{N}{(N-1)} \dfrac{[C_\nu(n)-C_\nu(n-1)]s_d^2}{C_\nu(n)};$

(III) $\qquad v_3(\bar{y}_\nu) = \dfrac{C_{\nu-1}(n-1)}{C_\nu(n)} s_d^2;$

(IV) $\qquad v_4(\bar{y}_\nu) = \left[\left(\dfrac{1}{\nu} - \dfrac{1}{N} \right) + \dfrac{(N-1)}{(N^n-N)} \right] s_d^2;$

(V) $\qquad v_5(\bar{y}_\nu) = \left[\left(\dfrac{1}{\nu} - \dfrac{1}{N} \right) + N^{1-n} \left(1 - \dfrac{1}{\nu} \right) \right] s_d^2 \quad$ (to be used for $\nu > 1$).

The estimate (II) is known to be uniformly better than (I). It appears difficult to give direct proofs of relative efficiencies of these estimators. The estimators (IV) and (V) were given by Des Raj and Khamis (1958). The estimator (V) is conditionally unbiased for $\nu > 1$. Des Raj and Khamis suggested the use of (V) for $\nu > 1$.

It is easy to see that

$$v_4 = v_5 \frac{N^n}{(N^n-N)}. \qquad \ldots \quad (6.1)$$

A little comparison will, now, show that the conditional variance of (V) is less than the variance of (IV). The amount of decrease in the variance is given by

$$V(v_4) - V(v_5/\nu > 1) = \frac{1}{N^{n-1}} E(v_4^2). \qquad \ldots \quad (6.2)$$

In general, this leads to the conclusion that any estimator $\hat{\sigma}^2$ of σ^2 which is unbiased for σ^2 and is equal to zero for $\nu = 1$, can be reduced to give a conditionally unbiased estimate of σ^2 for $\nu > 1$ whose conditional variance will be less than the variance of $\hat{\sigma}^2$. This conditionally improved estimator is related with $\hat{\sigma}^2$ by the following equation.

$$\hat{\sigma}^2 = \hat{\sigma}_{im}^2 \left[\frac{N^n}{(N^n-N)} \right] \qquad \ldots \quad (6.3)$$

where σ_{im}^2 stands for the conditionally improved estimator of σ^2.

Numerical example. To study the relative efficiency of the estimators of $V(\bar{y}_\nu)$, we consider the following three populations given by Yates and Grundy (1953).

TABLE 1. THREE POPULATIONS GIVEN BY
YATES AND GRUNDY

population	A	B	C
unit	Y_j	Y_j	Y_j
1	0.5	0.8	0.2
2	1.2	1.4	0.6
3	2.1	1.8	0.9
4	3.2	2.0	0.8
ΣY_j	7.0	6.0	2.5

These populations were deliberately chosen by them as being more extreme than will be normally encountered in practice.

The table below gives variances of unbiased estimators of $V(\bar{y}_\nu)$ when $n = 3$. $V(v_1)$ is not given as $V(v_1) > V(v_3)$.

TABLE 2. VARIANCES OF UNBIASED ESTIMATORS OF $V(Y)$

population	$V(Y_\nu)$	$V(v_2)$	$V(v_3)$	$V(v_4)$	$V(v_5\|\nu>1)$
A	0.29823	0.04940	0.05222	0.09017	0.07897
B	0.06125	0.00220	0.00232	0.00396	0.00348
C	0.020964	0.000279	0.000293	0.000490	0.000432

The results show that for the three populations

$$V(v_2) < V(v_3) < V(v_5|v>|) < V(v_4). \qquad \dots \quad (6.4)$$

Thus v_2 appears to be most efficient estimator of $V(\bar{y}_\nu)$.

For $n = 2$, v_2 and v_3 are identical. The comparison thus strongly suggests the use of v_2 for estimating $V(\bar{y}_\nu)$.

For getting estimators of $V(t)$, where t is any unbiased estimator of $\bar{Y}$, the following procedure may be adopted.

$$v(t) = t^2 - \text{est}\ (\bar{Y}^2), \qquad \dots \quad (6.5)$$

where est $(\bar{Y}^2)$ stands for an unbiased estimator of $\bar{Y}^2$ and can be obtained from any of the relations

$$\text{est}\ (\bar{Y}^2) = v_i(\bar{y}_\nu) - \bar{y}_\nu^2 \qquad (i = 1, 2, 3, 4, 5). \qquad \dots \quad (6.6)$$

From the example considered, it is expected that

$$\text{est}\ (\bar{Y}^2) = v_i(\bar{y}_\nu) - \bar{y}_\nu^2 \qquad (i = 2, 3) \qquad \dots \quad (6.7)$$

would fare better than the remaining estimators of $\bar{Y}^2$.

TABLE 3. VALUES OF $\dfrac{C_m(n-1)}{C_m(n)}$

$n \rightarrow$ m	2	3	4	5	6	7	8	9
1				1.000000 for all n				
2	0	.3333333	.4285714	.4666667	.4838710	.4920635	.4960630	.4980392
3		0	.1666667	.2400000	.2777778	.2990033	.3115942	.3193388
4			0	.1000000	.1538462	.1857143	.2057613	.2189189
5				0	.0666667	.1071429	.1333333	.1510574
6					0	.0476190	.0789474	.1005291
7						0	.0357143	.0606061
8							0	.0277778
9								0

$n \rightarrow$ m	10	11	12	13	14	15	16	17
1				1.000000 for all n				
2	.4990215	.4995112	.4997557	.4998779	.4999390	.4999695	.4999847	.4999924
3	.3242229	.3273569	.3293923	.3307253	.3316032	.3321838	.3325687	.3328243
4	.2278258	.2339966	.2383479	.2414585	.2437059	.2453432	.2465438	.2474286
5	.1634568	.1723544	.1788676	.1837118	.1873611	.1901391	.1922722	.1939216
6	.1159154	.1271791	.1355998	.1420028	.1469395	.1507901	.1538225	.1562302
7	.0785714	.0918937	.1019882	.1097724	.1158627	.1206858	.1245448	.1276595
8	.0480000	.0631313	.0747043	.0837155	.0908370	.0965357	.1011446	.1049062
9	.0222222	.0389610	.0518519	.0619607	.0700084	.0764970	.0817858	.0861370
10	0	.0181818	.0322581	.0433566	.0522416	.0594466	.0653539	.0702435
11		0	.0151515	.0271493	.0367965	.0446549	.0511277	.0565107
12			0	.0128205	.0231660	.0316239	.0386164	.0444536
13				0	.0109890	.0200000	.0274725	.0337299
14					0	.0095238	.0174419	.0240896
15						0	.0083333	.0153453
16							0	.0073529
17								0

TABLE 3.   VALUES OF $\dfrac{C_m(n-1)}{C_m(n)}$ (Contd.)

$n \rightarrow$ m	18	19	20	21	22	23	24	25
1				1.000000 for all n				
2	.4999962	.4999981	.4999990	.4999995	.4999998	.4999999	.4999999	.5000000
3	.3329943	.3331075	.3331828	.3332330	.3332665	.3332888	.3333036	.3333135
4	.2480833	.2485693	.2489308	.2492003	.2494015	.2495518	.2496643	.2497484
5	.1952045	.1962073	.1969942	.1976139	.1981031	.1984904	.1987974	.1990412
6	.1581551	.1597031	.1609542	.1619699	.1627974	.1634737	.1640281	.1644835
7	.1301924	.1322655	.1339720	.1353836	.1365562	.1375340	.1383520	.1390384
8	.1080005	.1105635	.1126991	.1144881	.1159936	.1172659	.1183450	.1192632
9	.0897460	.0927608	.0952950	.0974369	.0992561	.1008081	.1021373	.1032797
10	.0743243	.0777549	.0806574	.0831272	.0852394	.0870540	.0886193	.0899745
11	.0610250	.0648389	.0680820	.0708560	.0732409	.0753009	.0770878	.0786436
12	.0493679	.0535361	.0570949	.0601513	.0627902	.0650794	.0670738	.0688180
13	.0390147	.0435116	.0473637	.0506832	.0535590	.0560624	.0582511	.0601724
14	.0297189	.0345220	.0386476	.0422127	.0453100	.0480141	.0503852	.0524727
15	.0212963	.0263854	.0307669	.0345620	.0378671	.0407595	.0433020	.0455460
16	.0136054	.0189618	.0235845	.0275957	.0310960	.0341657	.0368697	.0392613
17	.0065359	.0121457	.0169935	.0212082	.0248926	.0281295	.0309861	.0335172
18	0	.0058480	.0109091	.0153161	.0191746	.0225696	.0255705	.0282338
19		0	.0052632	.0098522	.0138756	.0174206	.0205584	.0233471
20			0	.0047619	.0089419	.0126294	.0158973	.0188053
21				0	.0043290	.0081522	.0115440	.0145657
22					0	.0039526	.0074627	.0105929
23						0	.0036232	.0068571
24							0	.0033333
25								0

$n \rightarrow$ m	26	27	28	29	30	31	32	33
1				1.0000000	for all n			
2				.5000000	for $n > 24$			
3	.3333201	.3333245	.3333275	.3333294	.3333307	.3333316	.3333322	.3333326
4	.2498115	.2498587	.2498940	.2499206	.2499404	.2499553	.2499665	.2499749
5	.1992352	.1993895	.1995125	.1996106	.1996889	.1997514	.1998012	.1998411
6	.1648585	.1651676	.1654230	.1656342	.1658090	.1659539	.1660741	.1661738
7	.1396157	.1401025	.1405137	.1408616	.1411565	.1414068	.1416195	.1418004
8	.1200468	.1207173	.1212923	.1217864	.1222119	.1225788	.1228958	.1231699
9	.1042646	.1051161	.1058542	.1064955	.1070539	.1075409	.1079665	.1083390
10	.0911518	.0921776	.0930737	.0938586	.0945476	.0951537	.0956879	.0961594
11	.0800030	.0811943	.0822415	.0831643	.0839795	.0847012	.0853413	.0859103
12	.0703490	.0716970	.0728875	.0739417	.0748775	.0757102	.0764525	.0771157
13	.0618648	.0633607	.0646867	.0658656	.0669162	.0678548	.0686950	.0694488
14	.0543172	.0559525	.0574068	.0587037	.0598633	.0609026	.0618362	.0626766
15	.0475339	.0493009	.0508763	.0522851	.0535481	.0546833	.0557058	.0566289
16	.0413846	.0432760	.0449662	.0464810	.0478422	.0490684	.0501756	.0511775
17	.0357685	.0377778	.0395768	.0411922	.0426467	.0439595	.0451473	.0462243
18	.0306065	.0327276	.0346298	.0363408	.0378840	.0392793	.0405439	.0416926
19	.0258351	.0280625	.0300629	.0318649	.0334925	.0349665	.0363044	.0375215
20	.0214030	.0237315	.0258255	.0277142	.0294224	.0309714	.0323793	.0336619
21	.0172679	.0196929	.0218762	.0238477	.0256329	.0272536	.0287284	.0300736
22	.0133950	.0159121	.0181807	.0202314	.0220902	.0237795	.0253185	.0267237
23	.0097547	.0123601	.0147103	.0168368	.0187662	.0205213	.0221217	.0235844
24	.0063224	.0090123	.0114409	.0136400	.0156371	.0174553	.0191148	.0206327
25	.0030769	.0058480	.0083516	.0106206	.0126827	.0145616	.0162777	.0178488
26	0	.0028490	.0054250	.0077611	.0098857	.0118229	.0135936	.0152159
27		0	.0026455	.0050463	.0072310	.0092245	.0110478	.0127193
28			0	.0024631	.0047059	.0067535	.0086276	.0103467
29				0	.0022989	.0043988	.0063218	.0080868
30					0	.0021505	.0041209	.0059303
31						0	.0020161	.0038685
32							0	.0018939
33								0

TABLE 3. VALUES OF $\dfrac{C_m(n-1)}{C_m(n)}$ (Contd.)

$n \rightarrow$ m	34	35	36	37	38	39	40	41	42
1				1.0000000	for	all n			
2				.5000000	for	$n>24$			
				.3333333	for	$n>38$			
3	.3333328	.3333330	.3333331	.3333332	.3333332	.3333333	.3333333 for $n>38$		
4	.2499812	.2499859	.2499894	.2499921	.2499940	.2499955	.2499966	.2499975	.2499981
5	.1998730	.1998984	.1999188	.1999350	.1999480	.1999584	.1999667	.1999734	.1999787
6	.1662566	.1663255	.1663827	.1664302	.1664698	.1665027	.1665301	.16615530	.1665720
7	.1419544	.1420856	.1421976	.1422931	.1423746	.1424442	.1425037	.1425546	.1425981
8	.1234073	.1236131	.1237916	.1239467	.1240815	.1241988	.1243008	.1243897	.1244671
9	.1086654	.1089519	.1092036	.1094250	.1096199	.1097916	.1099431	.1100767	.1101948
10	.0965764	.0969451	.0972731	.0975637	.0978221	.0980519	.0982566	.0984390	.0986017
11	.0864167	.0868683	.0872715	.0876321	.0879548	.0882441	.0885036	.0887367	.0889462
12	.0777093	.0782413	.0787190	.0791485	.0795352	.0798837	.0801982	.0804824	.0807393
13	.0701262	.0707361	.0712861	.0717828	.0722320	.0726388	.0730075	.0733423	.0736464
14	.0634345	.0641193	.0647390	.0653007	.0658105	.0662739	.0666957	.0670799	.0674305
15	.0574637	.0582202	.0589068	.0595311	.0600994	.0606176	.0610907	.0615232	.0619189
16	.0520858	.0529109	.0536617	.0543460	.0549707	.0555417	.0560643	.0565434	.0569830
17	.0472028	.0480935	.0489058	.0496476	.0503263	.0509481	.0515186	.0520426	.0525245
18	.0427382	.0436916	.0445626	.0453597	.0460903	.0467608	.0473773	.0479446	.0484674
19	.0386311	.0396445	.0405718	.0414218	.0422021	.0429196	.0435802	.0441893	.0447515
20	.0348327	.0359036	.0368848	.0377855	.0386136	.0393761	.0400792	.0407284	.0413286
21	.0313031	.0324290	.0334620	.0344113	.0352853	.0360910	.0368350	.0375228	.0381596
22	.0280093	.0291880	.0302707	.0312668	.0321848	.0330321	.0338154	.0345404	.0352124
23	.0249241	.0261535	.0272838	.0283248	.0292852	.0301725	.0309937	.0317545	.0324605
24	.0220242	.0233023	.0244785	.0255627	.0265639	.0274898	.0283474	.0291428	.0298815
25	.0192902	.0206152	.0218354	.0229613	.0240017	.0249648	.0258575	.0266863	.0274566
26	.0167052	.0180754	.0193382	.0205041	.0215825	.0225813	.0235080	.0243689	.0251697
27	.0142550	.0156687	.0169725	.0181771	.0192920	.0203255	.0212849	.0221769	.0230072
28	.0119270	.0133827	.0147261	.0159681	.0171183	.0181852	.0191763	.0200983	.0209572
29	.0097103	.0112066	.0125883	.0138664	.0150508	.0161500	.0171717	.0181228	.0190092
30	.0075954	.0091310	.0105497	.0118628	.0130802	.0142107	.0152620	.0162412	.0171545
31	.0055740	.0071475	.0086021	.0099490	.0111984	.0123592	.0134393	.0144458	.0153849
32	.0036386	.0052489	.0067382	.0081179	.0093983	.0105885	.0116964	.0127294	.0136937
33	.0017825	.0034286	.0049515	.0063630	.0076735	.0088922	.0100272	.0110858	.0120745
34	0	.0016807	.0032362	.0046786	.0060183	.0072647	.0084259	.0095094	.0105218
35		0	.0015873	.0030597	.0044277	.0057010	.0068877	.0079955	.0090309
36			0	.0015015	.0028972	.0041965	.0054081	.0065394	.0075972
37				0	.0014225	.0027473	.0039829	.0051372	.0062168
38					0	.0013495	.0026087	.0037853	.0048862
39						0	.0012821	.0024804	.0036020
40							0	.0012195	.0023613
41								0	.0011614
42									0

TABLE 3. VALUES OF $\frac{C_m(n-1)}{C_m(n)}$ (Contd.)

$n \rightarrow$ m	43	44	45	46	47	48	49	50
1				1.0000000	for	all	n	
2				.5000000	for	$n > 24$		
3				.3333333	for	$n > 38$		
4	.2499986	.2499989	.2499992	.2499994	.2499996	.2499997	.2499997	.2499998
5	.1999830	.1999864	.1999891	.1999913	.1999930	.1999944	.1999955	.1999964
6	.1665878	.1666009	.1666119	.1666210	.1666287	.1666350	.1666403	.1666447
7	.1426354	.1426672	.1426945	.1427178	.1427378	.1427549	.1427695	.1427821
8	.1245346	.1245935	.1246897	.1246897	.1247288	.1247629	.1247928	.1248188
9	.1102991	.1103913	.1104729	.1105451	.1106090	.1106656	.1107158	.1107602
10	.0987470	.0988767	.0989927	.0990965	.0991893	.0992724	.0993468	.0994135
11	.0891346	.0893043	.0894572	.0895950	.0897194	.0898316	.0899330	.0900246
12	.0809719	.0811826	.0813736	.0815469	.0817042	.0818471	.0819771	.0820953
13	.0739230	.0741749	.0744043	.0746136	.0748045	.0749789	.0751383	.0752840
14	.0677505	.0680431	.0683107	.0685558	.0687804	.0689864	.0691754	.0693490
15	.0622814	.0626139	.0629191	.0631995	.0634574	.0636947	.0639132	.0641147
16	.0573868	.0577581	.0581000	.0584150	.0587054	.0589736	.0592212	.0594502
17	.0529683	.0533774	.0537549	.0541035	.0544259	.0547241	.0550003	.0552563
18	.0489498	.0493954	.0498074	.0501888	.0505420	.0508696	.0511736	.0514560
19	.0452712	.0457520	.0461974	.0466104	.0469937	.0473498	.0476809	.0479890
20	.0418842	.0423991	.0428768	.0433204	.0437328	.0441165	.0444738	.0448069
21	.0387498	.0392976	.0398064	.0402796	.0407202	.0411307	.0415135	.0418708
22	.0358361	.0364155	.0369545	.0374563	.0379240	.0383604	.0387679	.0391488
23	.0331164	.0337264	.0342944	.0348239	.0353180	.0357794	.0362108	.0366145
24	.0305684	.0312080	.0318041	.0323603	.0328798	.0333656	.0338201	.0342459
25	.0281735	.0288416	.0294649	.0300469	.0305911	.0311003	.0315773	.0320245
26	.0259157	.0266113	.0272608	.0278679	.0284359	.0289679	.0294666	.0299345
27	.0237812	.0245035	.0251784	.0258097	.0264008	.0269548	.0274745	.0279626
28	.0217583	.0225064	.0232059	.0238606	.0244740	.0250494	.0255895	.0260971
29	.0198366	.0206097	.0213330	.0220104	.0226455	.0232416	.0238016	.0243281
30	.0180073	.0188046	.0195510	.0202504	.0209065	.0215227	.0221019	.0226468
31	.0162624	.0170832	.0178520	.0185728	.0192493	.0198850	.0204828	.0210456
32	.0145950	.0154386	.0162291	.0169707	.0176670	.0183216	.0189376	.0195177
33	.0129991	.0138648	.0146764	.0154381	.0161537	.0168267	.0174603	.0180573
34	.0114690	.0123563	.0131884	.0139696	.0147040	.0153949	.0160456	.0166590
35	.0100000	.0109082	.0117602	.0125605	.0133131	.0140214	.0146887	.0153181
36	.0085877	.0095162	.0103876	.0112065	.0119767	.0127020	.0133856	.0140305
37	.0072281	.0081762	.0090667	.0099036	.0106911	.0114329	.0121323	.0127924
38	.0059176	.0068852	.0077939	.0086484	.0094527	.0102106	.0109254	.0116003
39	.0046531	.0056395	.0065662	.0074378	.0082585	.0090321	.0097619	.0104512
40	.0034317	.0044364	.0053806	.0062689	.0071056	.0078944	.0086390	.0093423
41	.0022506	.0032731	.0042344	.0051391	.0059913	.0067951	.0075540	.0082710
42	.0011074	.0021475	.0031254	.0040460	.0049135	.0057319	.0065047	.0072351
43	0	.0010571	.0020513	.0029874	.0038698	.0047024	.0054889	.0062324
44		0	.0010101	.0019614	.0028584	.0037049	.0045047	.0052611
45			0	.0009662	.0018773	.0027375	.0035504	.0043192
46				0	.0009251	.0017986	.0026242	.0034053
47					0	.0008865	.0017246	.0025177
48						0	.0008503	.0016552
49							0	.0008163
50								0

Note : Values of $\frac{C_{m-1}(n-1)}{C_m(n)}$ can also be obtained from $\frac{C_m(m-1)}{C_m(n)}$ by using following relation

$$\frac{1}{m} = \frac{C_m(n-1)}{C_m(n)} + \frac{C_{m-1}(n-1)}{C_m(n)} .$$

7. COMPARISON BETWEEN WITH AND WITHOUT REPLACEMENT SIMPLE RANDOM SAMPLING SCHEMES

In conclusion let us compare the two simple random sampling schemes for the purpose of estimation of $\bar{Y}$. If we draw a simple random sample with replacement of size n, then the variance of the sample mean is σ^2/n. Further, in a simple random sample without replacement of size n, the variance of the sample mean is $\frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$. Since

$$\frac{\sigma^2}{n} \cdot \frac{(N-n)}{(N-1)} < \frac{\sigma^2}{n}$$

it is usually claimed that sampling without replacement is better than sampling with replacement. Basu (1958) has pointed out that this comparison is not fair because the cost of selecting a sample of size n in sampling without replacement is greater than the cost of selecting a sample in sampling with replacement. For comparing the two sampling schemes, it would be appropriate to take into account the cost involved in the selection of two different samples. The comparison, thus, mainly depends on the choice of the cost function, and no sampling scheme can be said to be superior to the other unless the cost function is known in advance. Let us, for illustration, consider the case where the cost of sampling is proportional to the number of distinct units drawn. Thus the expected cost of selecting a sample with replacement of size n is equivalent to the cost of selecting a sample without replacement of size $E(\nu) = N\left[1 - \left(\frac{N-1}{N}\right)^n\right]$. Basu has shown that in this situation the sample mean of the sample with replacement is worse than the sample mean of the equivalent sample without replacement. We now compare the sample mean $\bar{y}$ of the equivalent sample without replacement with the following estimator of with replacement sample :

$$\bar{y}_{\nu(2)} = \frac{[N\nu/(N-\nu)]}{E[N\nu/(N-\nu)]}\bar{y}_\nu.$$

It has been shown that

$$V(\bar{y}_{\nu(2)}) = \frac{S^2}{E[N\nu/(N-\nu)]} + \bar{Y}^2 V\left[\frac{N\nu/(N-\nu)}{E\{N\nu/(N-\nu)\}}\right], \qquad \ldots \ (7.1)$$

and

$$V(\bar{y}) = \left[\frac{1}{E(\nu)} - \frac{1}{N}\right]S^2. \qquad \ldots \ (7.2)$$

Since $N\nu/(N-\nu)$ is a convex function of $\nu (1 \leqslant \nu \leqslant n < N)$,

$$E\left[N\nu/(N-\nu)\right] \geqslant NE\nu/[N-E\nu] = \left[\frac{1}{E\nu} - \frac{1}{N}\right]^{-1} \qquad \ldots \ (7.3)$$

From (7.3), it is evident that the first component of $V(\bar{y}_{\nu(2)})$ is smaller than $V(\bar{y})$. Thus for a population whose coefficient of variation is sufficiently large $V(\bar{y}_{\nu(2)})$ would be smaller than $V(\bar{y})$. This comparison shows that the sample mean of without replacement sample cannot be uniformly better than all estimators of with replacement sampling.

However, the comparison made above is not very satisfactory. First, because of the linearity of the cost function and secondly, because $E(\nu)$ is not necessarily an integer. We hope that for some other cost functions also, similar situations may be found out where with replacement sampling would fare better than without replacement sampling.

ACKNOWLEDGEMENT

REFERENCES

BASU, D. (1958) : On sampling with and without replacement. *Sankhyā*, **20**, 287-294.

DES RAJ and KHAMIS, S. H. (1958) : Some remarks on sampling with replacement. *Ann. Math. Stat.*, **39**, 550-557.

FRASER, D. A. S. (1957) : *Non-parametric Methods in Statistics*, John Wiley and Sons, New York.

GUPTA, H. (1950) : Tables of distributions. *Research Bulletin of the East Punjab University*, No. 2, 3. 13-44.

HARTLEY, H. O. and ROSS, A. (1954) : Unbiased ratio estimators. *Nature*, **174**, 270-271.

HORVITZ, D. G. and THOMPSON, D. J. (1952) : A generalization of sampling without replacement from finite universes. *J. Amer. Stat. Ass.*, **47**, 663-685.

MURTHY, M. N. (1961) : Some problems of estimation in sampling from finite populations. Thesis submitted for the Ph.D degree of the Indian Statistical Institute, 105-112.

PATHAK, P. K. (1961) : On the evaluation of moments of distinct units in a sample, *Sankhyā*, Series A, **28**, 415-420.

YATES, F. and GRUNDY, P. M. (1953) : Selection without replacement from within strata with probability proportional to size. *J. Roy. Stat. Soc.*, Series B, **15**, 253-261.

ALMOST UNBIASED ESTIMATORS BASED ON
INTERPENETRATING SUB-SAMPLES

By M. N. MURTHY
Indian Statistical Institute

SUMMARY. In this paper a technique is given for estimating unbiasedly any non-linear function of estimable parameters. The technique consists in estimating the bias of the usual estimator using estimates based on interpenetrating sub-samples and then correcting the estimator for its bias.

1. INTRODUCTION

The question of evolving a generalized unbiased estimator for any sample design has been considered by Midzuno (1950), Godambe (1955) and Nanjamma, Murthy and Sethi (1959) for certain classes of parameters. Murthy (1962) has suggested a technique of generating unbiased estimators for any sample design for the class of parameters which can be expressed as a sum of single-valued set functions defined over a class of sets of units belonging to the finite population under consideration. Examples of such parameters are the population total Y and the population variance which can be expressed respectively as

$$Y = \sum_{i=1}^{N} Y_i$$

and

$$\sigma^2 = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j>i} (Y_i - Y_j)^2.$$

In this paper we shall supplement the generalized theory of unbiased estimation (Murthy, 1962) by giving a procedure of obtaining unbiased (or almost unbiased) estimators for non-linear functions of parameters each of which can be expressed as single-valued set function defined over a class of sets of units belonging to a finite population. Examples of such parameters are given by ratio of population totals of two characteristics, population standard deviation, correlation coefficient, etc.

The procedure of obtaining this unbiased estimator consists in estimating the bias of the usual estimator which is taken as the same non-linear function of unbiased estimators of the parameters as the parametric function under consideration, on the basis of interpenetrating sub-sample estimates. This procedure is based on the technique used by Murthy and Nanjamma (1959) in estimating the bias of a ratio estimator.

The procedure given in this paper is likely to be of much help in survey practice, since the estimation of relationships between characteristics and between parameters, such as a ratio of population totals of two characteristics or the population coefficient of variation are usually of much interest in sample surveys.

2. Parametric function

Let the parametric function $f(\theta)$ be a single valued non-linear function of the parameters $(\theta_1, \theta_2, ..., \theta_k)$, where θ_i $(i = 1, 2, ..., k)$ can be expressed as

$$\theta_i = \sum_{a_i \epsilon A_i} f_i(a_i) \qquad \qquad ... \quad (2.1)$$

where $f_i(a_i)$ is a single-valued set function defined over the class 'A_i' of sets 'a_i' consisting of units belonging to the population X.

Suppose we have defined the sample space 'S' of samples 's' with a suitable probability measure such that it is possible to estimate the parameters $(\theta_1, \theta_2, ..., \theta_k)$ unbiasedly using the procedure given by Murthy (1962). That is, it is assumed that the sample space is so specified that each $a_i \epsilon A_i$ $(i = 1, 2, ..., k)$ occurs in at least one 's' and that each 's' contains at least one set 'a_i' in 'A_i' $(i = 1, 2, ..., k)$. Then a generalized unbiased estimator of θ_i $(i = 1, 2, ..., k)$ is given by

$$t_i = \hat{\theta}_i = \sum_{a_i \subset s} f_i(a_i)\phi_i(s, a_i)/P(s) \qquad \qquad ... \quad (2.2)$$

where

$$\sum_{s \supset a_i} \phi_i(s, a_i) = 1.$$

In fact, we can make the above formulation more general by relaxing the assumption that θ_i's $(i = 1, 2, ..., k)$ are estimated from the same samples. In other words, θ_i $(i = 1, 2, ..., k)$ may be estimated on the basis of the same, overlapping or non-overlapping samples drawn with the same or different sample designs.

Let $(t_1, t_2, ..., t_k)$ be unbiased estimators of the parameters $(\theta_1, \theta_2, ..., \theta_k)$. Then an estimator of $f(\theta)$ can be taken as $f(t)$. If $f(\theta)$ is a linear function, obviously $f(t)$ will be unbiased for $f(\theta)$. But here we are taking $f(\theta)$ as a non-linear function of $(\theta_1, \theta_2, ..., \theta_k)$ and hence $f(t)$ will, in general, be biased for $f(\theta)$.

3. Bias and mean square error

In this section approximate expressions for the bias and the mean square error of the estimator of $f(t)$ are obtained by using Taylor series symbolically. It may be noted that in statistical practice one is interested not so much in the convergence properties of the infinite series representing a function, but in finding out whether the first few terms of that series will give a good approximation to the function. Because of this, the question of the validity of the application of Taylor series expansion to the case of a finite population estimator will not be considered here. However, it will be assumed that the estimator t_i is such that $\left| \dfrac{t_i - \theta_i}{\theta_i} \right| < 1$, especially for estimators occurring in the denominator of the function $f(t)$ so that the first few terms of the expansion can be expected to give a good approximation to the function. This latter statement has been empirically verified in the case of applying this expansion to a ratio estimator.

If the sample size is fairly large, the assumption $\left|\dfrac{t_i-\theta_i}{\theta_i}\right| < 1$ will be valid. Let $t_i=\theta_i\,(1+e_i)$, $(i=1, 2, ..., k)$ and $t=(t_1, t_2, ..., t_k)$, $=(\theta_1, \theta_2, ..., \theta_k)$, $e=\theta(e_1, e_2, ...,e_k)$. Expanding $f(t)$ in a Taylor series about $t=\theta$ and neglecting terms of degree greater than 2 in e's, we get

$$f(t) = f(\theta)+\sum_{i=1}^{k} \theta_i e_i \left(\frac{\partial f}{\partial t_i}\right)_{t=\theta}$$

$$+\frac{1}{2}\left[\sum_{i=1}^{k} \theta_i^2 e_i^2\left(\frac{\partial^2 f}{\partial t_i^2}\right)_{t=\theta} +2\sum_{i=1}^{k}\sum_{j>i} \theta_i\theta_j e_i e_j\left(\frac{\partial^2 f}{\partial t_i\partial t_j}\right)_{t=\theta}\right] \quad ... \quad (3.1)$$

It may be observed that for certain parameters there will be no terms of degree greater than 2 to neglect. An example of such a parameter is the product $\theta_1\theta_2$ with the estimator $t_1 t_2$. Taking expected value of $f(t)$ in (3.1), we find that the bias of $f(t)$ correct to the second degree of approximation is given by

$$B[f(t)] =\frac{1}{2}\left[\sum_{i=1}^{k}\left(\frac{\partial^2 f}{\partial t_i^2}\right)_{t=\theta} \mu_2(ii)+2\sum_{i=1}^{k}\sum_{j>i}\left(\frac{\partial^2 f}{\partial t_i\partial t_j}\right)_{t=\theta}\mu_2(ij)\right] \quad ... \quad (3.2)$$

where $\qquad \mu_2(ij) = E(t_i-\theta_i)(t_j-\theta_j)$, $\quad i, j = 1, 2, ..., k$.

The mean square error of $f(t)$ to the second degree of approximation is given by

$$M[f(t)] = E[f(t)-f(\theta)]^2=E\left[\sum_{i=1}^{k}\theta_i e_i\left(\frac{\partial f}{\partial t_i}\right)_{t=\theta}\right]$$

$$= \sum_{i=1}^{k}\left(\frac{\partial f}{\partial t_i}\right)_{t=\theta}^2 \mu_2(ii)+2\sum_{i=1}^{k}\sum_{j>i}\left(\frac{\partial f}{\partial t_i}\right)_{t=\theta}\left(\frac{\partial f}{\partial t_j}\right)_{t=\theta}\mu_2(ij). \quad ... \quad (3.3)$$

4. Biases of two estimators

Suppose the sample on which the estimate t_i of θ_i $(i = 1, 2, ..., k)$ is based is selected in the form of n independent interpenetrating sub-samples. Let t_{is} be the unbiased estimate of θ_i based on the s-th independent interpenetrating sub-sample $(i = 1, 2, ..., k;\ s = 1, 2, ..., n)$. In this case let us consider the following two estimators T_1 and T_n of $f(\theta)$.

$$T_1 = \frac{1}{n}\sum_{s=1}^{n} f(t_s) \qquad ... \quad (4.1)$$

where $\qquad t_s = (t_{1s}, t_{2s}, ..., t_{ks})$, $\quad (s = 1, 2, ..., n)$,

and $\qquad T_n = f(\bar{t}) \qquad ... \quad (4.2)$

$$\bar{t} = (\bar{t}_1, \bar{t}_2, ..., \bar{t}_k), \quad \bar{t}_i = \frac{1}{n}\sum_{s=1}^{n} t_{is}, \quad (i = 1, 2, ..., k).$$

Applying the result (3.2) to T_n in (4.2) we get

$$B(T_n) = \frac{1}{2} \left[\sum_{i=1}^{k} \left(\frac{\partial^2 f}{\partial t_i^2} \right)_{t=\theta} \bar{\mu}_2(ii) + 2 \sum_{i=1}^{k} \sum_{j>i} \left(\frac{\partial^2 f}{\partial t_i \partial t_j} \right)_{t=\theta} \bar{\mu}_2(ij) \right]$$

where

$$\bar{\mu}_2(ij) = E(\bar{t}_i - \theta_i)(\bar{t}_j - \theta_j) = \frac{1}{n^2} \sum_{s=1}^{n} \mu_{2s}(ij),$$

$$\mu_{2s}(ij) = E(t_{is} - \theta_i)(t_{js} - \theta_j).$$

That is

$$B_n = B(T_n) = \frac{1}{n^2} \sum_{s=1}^{n} \frac{1}{2} \left[\sum_{i=1}^{k} \left(\frac{\partial^2 f}{\partial t_i^2} \right)_{t=\theta} \mu_{2s}(ii) + 2 \sum_{i=1}^{k} \sum_{j>i} \left(\frac{\partial^2 f}{\partial t_i \partial t_j} \right)_{t=\theta} \mu_{2s}(ij) \right]$$

$$= \frac{1}{n^2} \sum_{s=1}^{n} B[f(t_s)]. \qquad \qquad \dots \ (4.3)$$

The bias of the estimator T_1 in (4.1) is given by

$$B_1 = B(T_1) = \frac{1}{n} \sum_{s=1}^{n} B[f(t_s)]. \qquad \dots \ (4.4)$$

Comparing (4.3) and (4.4) we find that the bias of the estimator T_1 is n times that of the estimator T_n.

5. Estimation of bias

As observed in Section 4, comparing the biases of the estimators T_1 and T_n, we get

$$B_1 = n \, B_n. \qquad \dots \ (5.1)$$

Using this result we can derive an unbiased estimator of the bias B_1.

$$E(T_1) = f(\theta) + B_1$$

$$E(T_n) = f(\theta) + B_n.$$

Hence

$$E(T_1 - T_n) = B_1 - B_n = (n-1)B_n.$$

Thus an unbiased estimator of B_n is given by

$$\hat{B}_n = \frac{T_1 - T_n}{n-1}. \qquad \dots \ (5.2)$$

The variance of the estimator of $\hat{B}_n$ is given by

$$V(\hat{B}_n) = \frac{V(T_1)}{(n-1)^2} (a^2 - 2\rho a + 1) \qquad \dots \ (5.3)$$

where $a^2 = V(T_n)/V(T_1)$, and ρ is the correlation coefficient between the estimators T_1 and T_n. For most of the sample designs a^2 and ρ will tend to 1 as the sample size increases and hence the variance of the bias estimator will tend to 0 as sample size increases. It may be observed that an unbiased estimator of the bias of T_1 is given by

$$\hat{B}_1 = \frac{n}{n-1} \ (T_1 - T_n). \qquad \qquad \dots \ (5.4)$$

6. (Almost) Unbiased Estimator

Since an unbiased estimator of the bias of the estimator T_n has been obtained in Section 5, the estimator T_n can be corrected for its bias, thereby obtaining an unbiased or almost unbiased estimator of $f(\theta)$ according as the third and higher degree terms in 'e' become 0 or not. In the latter case, the estimator is said to be almost unbiased since it is unbiased only to the second degree of approximation. The estimator corrected for its bias is given by

$$T_c = T_n - B_n = T_n - \frac{T_1 - T_n}{(n-1)} = \frac{nT_n - T_1}{(n-1)}. \qquad \dots \ (6.1)$$

It may be noted that this is the corrected estimator we get, even if we correct the estimator T_1 for its bias.

The variance of the corrected estimator is

$$V(T_c) = \frac{V(T_1)}{(n-1)^2} \ (n^2 a^2 - 2\rho a + 1). \qquad \qquad \dots \ (6.2)$$

The gain in precision in using T_c instead of T_n is given by

$$G(T_c) = \frac{M_n - V(T_c)}{M_n} = 1 - \frac{n^2 a^2 - 2n\rho a + 1}{(n-1)^2(a^2 + z^2)} \qquad \dots \ (6.3)$$

where z^2 is the ratio of the square of the bias of T_1 to the variance of T_1. If the subsample size is large z^2 will be negligibly small. Neglecting z^2 in the above expression, we find that the gain in precision will be positive if

$$(2n-1)a^2 - 2n\rho a + 1 < 0$$

which will be true if 'a' lies between the roots of the equation

$$(2n-1)a^2 - 2n\rho a + 1 = 0. \qquad \qquad \dots \ (6.4)$$

For given values of a and ρ, the minimum value of n which makes the corrected estimator more efficient and the value of n which maximises the gain are respectively given by

$$\left[\frac{(1-a^2)}{2a(\rho - a)} \right] + 1 \qquad \qquad \dots \ (6.5)$$

and

$$\frac{(1-\rho a)}{a(\rho - a)} \qquad \qquad \dots \ (6.6)$$

TABLE 1. SHOWING THE MINIMUM AND MAXIMUM VALUES
OF $G(T_c)$ AND THE CORRESPONDING VALUES OF n
FOR DIFFERENT VALUES OF ρ AND a $(\rho > a)$

sr. no.	a	ρ	minimum		maximum	
			n	$G(T_c)$	n	$G(T_c)$
1	0.6	0.7	6	0.0089	10	0.0192
2		0.8	3	0.0556	4	0.0988
3		0.9	2	0.0889	3	0.3056
4	0.7	0.8	4	0.0113	7	0.0266
5		0.9	2	0.1020	3	0.1684
6	0.8	0.9	3	0.0469	4	0.0486

Source : Murthy, M. N. and Nanjamma, N. S. (1959) : Almost
unbiased ratio estimates based on interpenetrating sub-sample esti-
mates, *Sankhyā*, **21**, 381-392.

7. ILLUSTRATIONS

In this section, the results derived in the previous sections are applied to
some particular cases.

Case (i) : $f(\theta) = \theta^k$. Let t be an unbiased estimator of θ based on any
sample design. Then an estimator of $f(\theta)$ is given by

$$f(t) = t^k. \qquad \qquad \text{.... (7.1)}$$

The bias and mean square error of $f(t)$ correct to the second degree of approximation
are given by

$$B[f(t)] = \frac{1}{2} k(k-1) C^2 f(\theta) \qquad \text{... (7.2)}$$

$$M[f(t)] = k^2 C^2 [f(\theta)]^2 \qquad \text{... (7.3)}$$

where C^2 is the relative variance of $t [= V(t)/\theta^2]$, since

$$\frac{df}{dt} = kt^{k-1} \text{ and } \frac{d^2 f}{dt^2} = k(k-1)t^{k-2}.$$

The bias relative to the mean square error is

$$\frac{B^2[f(t)]}{M[f(t)]} = \frac{1}{4} (k-1)^2 C^2. \qquad \text{... (7.4)}$$

From (7.2) and (7.4) we see that the bias of $f(t)$ and its contribution to the mean square
error both decrease as the sample size increases, since for most sample designs C^2
decreases with increase in sample size.

If t_s $(s = 1; 2, ..., n)$ are unbiased estimates of θ based on n independent interpenetrating sub-samples, the following two estimators T_1 and T_n of $f(\theta)$ can be considered.

$$T_1 = \frac{1}{n} \sum_{s=1}^{n} t_s^k \qquad \qquad \ldots \text{ (7.5)}$$

and

$$T_n = \bar{t}^k, \left(\bar{t} = \frac{1}{n} \sum_{s=1}^{n} t_s \right). \qquad \ldots \text{ (7.6)}$$

We have seen that the bias of T_1 is n times that of the bias of T_n. Hence an unbiased estimator of the bias of T_n is given by

$$\hat{B}(T_n) = \frac{\sum_{s=1}^{n} t_s^k - n\bar{t}^k}{n(n-1)} \qquad \ldots \text{ (7.7)}$$

and the corrected estimator is given by

$$T_c = \frac{n^2 \bar{t}^k - \sum_{s=1}^{n} t_s^k}{n(n-1)} \qquad \ldots \text{ (7.8)}$$

It may be noted that the expression for bias and the corrected estimator will be completely unbiased if k in $f(\theta)$ is 2.

Case (ii) : Correlation Coefficient (ρ). The correlation coefficient between two characteristics x and y is

$$\rho = \frac{\text{cov } (x, y)}{\sqrt{V(x)\, V(y)}}. \qquad \ldots \text{ (7.9)}$$

In this case the parametric function is of the form

$$f(\theta) = \frac{\theta_1}{\sqrt{\theta_2 \theta_3}}. \qquad \ldots \text{ (7.10)}$$

and the estimator is given by $\qquad f(t) = \frac{t_1}{\sqrt{t_2 t_3}} \qquad \ldots \text{ (7.11)}$

where t_1, t_2 and t_3 are unbiased estimators of θ_1, θ_2 and θ_3 respectively. The bias and mean square error of $f(t)$ correct to the second degree of approximation are given by

$$B[\,f(t)] = \frac{f(\theta)}{8} [3(v_{22} + v_{33}) - 4(v_{12} + v_{13}) + 2v_{23}] \qquad \ldots \text{ (7.12)}$$

and

$$M[f(t)] = \frac{[f(\theta)]^2}{4} [4v_{11} + (v_{22} + v_{33}) - 4(v_{12} + v_{13}) + 2v_{23}] \qquad \ldots \text{ (7.13)}$$

where

$$v_{ij} = \frac{E(t_i - \theta_i)(t_j - \theta_j)}{\theta_i \theta_j}.$$

Let t_{is} $(i = 1, 2, 3)$ be unbiased estimates based on the s-th independent interpenetrating sub-sample $(s = 1, 2, ..., n)$. Then using the two estimators

$$T_1 = \frac{1}{n} \sum_{s=1}^{n} \frac{t_{1s}}{\sqrt{t_{2s} t_{3s}}} \qquad \qquad ... \ (7.14)$$

$$T_n = \frac{\bar{t}_1}{\sqrt{\bar{t}_2 \bar{t}_3}} \qquad \qquad ... \ (7.15)$$

we get the following corrected estimator of ρ

$$T_c = \frac{n T_n - T_1}{(n-1)}. \qquad \qquad ... \ (7.16)$$

Case (iii) : Regression Estimator. Let y and x be unbiased estimators of the population totals Y and X respectively and let b be a consistent estimator of the regression coefficient obtained by taking the ratio of unbiased estimators of the covariance between x and y and the variance of x.

The regression estimator is $\quad \hat{y} = y + b(X - x)$. $\qquad \qquad ... \ (7.17)$

The estimator in this case is of the form

$$f(t) = t_1 + \frac{t_2}{t_3}(X - t_4). \qquad \qquad ... \ (7.18)$$

The bias and the mean square error of this estimator, correct to the second degree of approximation, are given by

$$B[f(t)] = \beta X(v_{34} - v_{24}) \qquad \qquad ... \ (7.19)$$

and $\qquad \qquad M[f(t)] = V(y) - 2\beta \text{ cov }(x, y) + \beta^2 V(x). \qquad ... \ (7.20)$

By defining the two estimators T_1 and T_n on the basis of n interpenetrating sub-sample estimates we get the corrected estimator as

$$T_c = \frac{n T_n - T_1}{(n-1)}.$$

Case (iv) : Skewness $(\beta_2 = \mu_4/\mu_2^2)$. The parametric function is of the form

$$f(\theta) = \theta_1/\theta_2^2$$

and an estimator of $f(\theta)$ is given by

$$f(t) = t_1/t_2^2$$

where t_1 and t_2 are unbiased estimators of θ_1 and θ_2 respectively. The bias and the mean square error of $f(t)$, correct to the second degree of approximation, are given by

$$B[f(t)] = \beta_2(3v_{22} - 2v_{12}) \qquad \qquad ... \ (7.21)$$

and $\qquad \qquad M[f(t)] = \beta_2^2(v_{11} + 4v_{22} - 4v_{12}) \qquad \qquad ... \ (7.22)$

where $\qquad \qquad v_{ij} = E(t_i - \theta_i)(t_j - \theta_j)/\theta_i \theta_j.$

Defining suitably the two estimators T_1 and T_n based on n interpenetrating sub-samples we get the corrected estimator, as before, as

$$\hat{\beta}_{20} = \frac{nT_n - T_1}{(n-1)}.$$

8. ESTIMATION OF BIAS

General Case. Suppose $f(\theta)$ is the parametric function of the parameters $(\theta_1, \theta_2, ..., \theta_k)$ and $f(t)$ is an estimator of $f(\theta)$ based on the estimators $(t_1, t_2, ..., t_k)$ which are unbiased for the parameters $(\theta_1, \theta_2, ..., \theta_k)$. Let $t_i = \theta_i + h_i$, $i = 1, 2, ..., k$. Applying Taylor series expansion to $f(t)$ about $t = \theta$ symbolically and neglecting terms of degrees greater than p in h_i's, we get

$$f(t) = f(\theta) + \sum_{j=1}^{p} \frac{1}{j!} \sum_{i_1, i_2, ..., i_j} (h_{i_1} h_{i_2} ..., h_{i_j}) \left(\frac{d^j f}{dt_{i_1} dt_{i_2} ... dt_{i_j}} \right)_{t=\theta} \quad ... \quad (8.1)$$

Taking the expected value of (8.1), we get the bias of $f(t)$ as

$$B[f(t)] = \sum_{j=2}^{p} \frac{1}{j!} \sum_{i_1, i_2, ..., i_j} \mu_j(i_1, i_2, ... i_j) \left[\frac{d^j f}{dt_{i_1} dt_{i_2} ... dt_{i_j}} \right]_{t=\theta} \quad ... \quad (8.2)$$

Suppose t_{is} is an unbiased estimate of θ_i based on the s-th independent inter-penetrating sub-sample ($i = 1, 2, ..., k$; $s = 1, 2, ... n$). Let us consider the following p estimators of $f(\theta)$

$$T_m = \frac{1}{\binom{n}{m}} \sum f(\bar{t}(m)), \quad m = 1, 2, ... p-1, n \quad ... \quad (8.3)$$

where $\qquad \bar{t}(m) = (\bar{t}_1(m), \bar{t}_2(m), ..., \bar{t}_k(m)),$

$\bar{t}_i(m)$ being the mean of the estimate t_i based on a combination of m sub-samples taken from the n independent interpenetrating sub-samples and Σ denotes summation over all combinations of m sub-samples formed out of n sub-samples.

The bias of T_m to the p-th degree of approximation is given by

$$B_m = B(T_m) = \frac{1}{\binom{n}{m}} \sum B[f(\bar{t}(m))]$$

$$= \frac{1}{\binom{n}{m}} \sum E \left[\sum_{j=2}^{p} \frac{1}{j!} \sum_{i_1, i_2, ..., i_j} (\bar{h}_{i_1} \bar{h}_{i_2} ... \bar{h}_{i_j}) \left(\frac{\partial^j f}{\partial t_{i_1} \partial t_{i_2} ... \partial t_{i_j}} \right) \right]_{t=\theta}$$

$$... \quad (8.4)$$

where $\bar{h}_{i_r} = \frac{1}{m} \sum\limits_{s=1}^{m} h_{i_r s}$. After simplification the bias of T_m may be expressed in the form

$$B_m = \sum_{j=2}^{p} \frac{A_j}{m^{j-1}} \quad (m = 1, 2, ..., p-1, n) \qquad \qquad ... \quad (8.5)$$

where A_j is a function of the j-th order moments and product moments of the estimators $(t_1, t_2, ..., t_k)$ and of terms of the form

$$\left[\frac{d^r f}{dt_{i_1} dt_{i_2} ... dt_{i_r}} \right]_{t=\theta}, \quad (r \geqslant j).$$

From (8.5) we see that in the series of estimation $\{T_m\}$, $B(T_{m+1}) < B(T_m)$.

Since

$$E(T_m) = f(\theta) + B_m,$$

we get

$$E(T_1 - T_m) = B_1 - B_m = \sum_{j=2}^{p} \left(1 - \frac{1}{m^{j-1}} \right) A_j. \qquad ... \quad (8.6)$$

Let $D_m = (T_1 - T_m)$. The equation (8.6) can be written as

$$E(D) = A \Lambda, \qquad \qquad ... \quad (8.7)$$

where

$$D = (D_2, D_3, ..., D_{p-1}, D_n)$$

$$A = (A_2, A_3, ..., A_{p-1}, A_p)$$

$$\Lambda = \begin{pmatrix} 1-\dfrac{1}{2} & 1-\dfrac{1}{3} & ... & 1-\dfrac{1}{p-1} & 1-\dfrac{1}{n-1} \\[2ex] 1-\dfrac{1}{2^2} & 1-\dfrac{1}{3^2} & ... & 1-\dfrac{1}{(p-1)^2} & 1-\dfrac{1}{(n-1)^2} \\[2ex] 1-\dfrac{1}{2^{p-1}} & 1-\dfrac{1}{3^{p-1}} & ... & 1-\dfrac{1}{(p-1)^{p-1}} & 1-\dfrac{1}{(n-1)^{p-1}} \end{pmatrix}.$$

It may be noted that in (8.7) we have $(p-1)$ equations in $(p-1)$ unknowns. It may be observed that we are considering p estimators since there are $(p-1)A$'s and $f(\theta)$ to be estimated. Solving (8.7) for A we get

$$A = E(D) \Lambda^{-1}. \qquad \qquad ... \quad (8.8)$$

Taking $B = (B_2, B_3, ..., B_{p-1}, B_n)$, we get

$$B = A(e + \Lambda) \qquad \qquad ... \quad (8.9)$$

where 'e' is a $(p-1, p-1)$ matrix whose elements are all equal to 1. Substituting in (8.9) the solution for A obtained in (8.8), we get unbiased estimators of the biases of the estimators, namely,

$$\hat{B} = D\Lambda^{-1}e - D.$$

That is

$$\hat{B}_m = \sum_j D_j \, S_{j-1} - D_m, \, (j = 2, 3, \ldots, p-1, n), \, (m = 2, 3, \ldots, p-1, n), \, \ldots \quad (8.10)$$

$S_{n-1} = S_{p-1}$, where S_j is the sum of the elements in the j-th row of Λ^{-1}.

Particular Cases : (i) $p = 2$. This is the case considered earlier. In this case, the following 2 estimators of $f(\theta)$ may be considered.

$$T_1 = \frac{1}{n} \sum_{s=1}^{n} f(t_s)$$

$$T_n = f(\bar{t}).$$

$$B_m = A/m, \, (m = 1, n).$$

Since

$$\Lambda = \left(1 - \frac{1}{n}\right), \quad \Lambda^{-1} = \frac{n}{n-1}, \quad S_1 = \frac{n}{n-1},$$

we get

$$\hat{B}_n = \frac{n}{n-1} \, (T_1 - T_n) - (T_1 - T_n)$$

$$= (T_1 - T_n)/(n-1).$$

Case (ii) : $p = 3$.

Let us consider the following three estimators of $f(\theta)$:

$$T_1 = \frac{1}{n} \sum_{s=1}^{n} f(t_s) \qquad \ldots \quad (8.11)$$

$$T_2 = \frac{2}{n(n-1)} \, \Sigma f(\bar{t}(2)) \qquad \ldots \quad (8.12)$$

$$T_n = f(\bar{t}). \qquad \ldots \quad (8.13)$$

$$B_m = \frac{A_2}{m} + \frac{A_3}{m^2}, \, (m = 1, 2, n). \qquad \ldots \quad (8.14)$$

Since

$$\Lambda = \begin{pmatrix} 1 - \dfrac{1}{2} & 1 - \dfrac{1}{n} \\[2mm] 1 - \dfrac{1}{2^2} & 1 - \dfrac{1}{n^2} \end{pmatrix}$$

and

$$\Lambda^{-1} = -\frac{4n^2}{(n-1)(n-2)} \begin{pmatrix} \dfrac{n^2-1}{n^2} & -\dfrac{n-1}{n} \\[2mm] -\dfrac{3}{4} & \dfrac{1}{2} \end{pmatrix},$$

we get after simplification

$$B_1 = \frac{n-2}{n-1}\, T_1 + \frac{4}{n-2}\, T_2 - \frac{n^2}{(n-1)(n-2)} T_n \qquad \dots \quad (8.15)$$

$$B_2 = -\frac{1}{n-1}\, T_1 + \frac{n+2}{n-2}\, T_2 - \frac{n^2}{(n-1)(n-2)}\, T_n \qquad \dots \quad (8.16)$$

$$B_n = -\frac{1}{n-1}\, T_1 + \frac{4}{n-2}\, T_2 - \frac{(3n-2)}{(n-1)(n-2)}\, T_{n'} \qquad \dots \quad (8.17)$$

REFERENCES

GODAMBE, V. P. (1955) : A unified theory of sampling from finite population. *J. Roy. Stat. Soc.*, 17(B), 269-278.

MIDZUNO, H. (1950) : An outline of the theory of sampling systems. *Ann. Math. Stat.*, 1, 149-156.

MURTHY, M. N. and NANJAMMA, N. S. (1959) : Almost unbiased estimates based on interpenetrating sub-sample estimates. *Sankhyā*, 21, 381-392.

MURTHY, M. N., NANJAMMA, N. S. and SETHI, V. K. (1959) : Some sampling systems providing unbiased ratio estimators. *Sankhyā*, 21, 299-314.

MURTHY, M. N. (1962) : Generalized unbiased estimation in sampling from finite populations. Submitted for publication.

Paper received : February, 1962.

ON SAMPLING WITH UNEQUAL PROBABILITIES

By P. K. PATHAK

Indian Statistical Institute

SUMMARY. This paper deals with the problem of deriving improved estimators in sampling schemes with unequal probabilities of selection. The improved estimator of the population total, Y, (Basu, 1958), is derived. In addition, two sets of estimators of Y and Y^2 are given. The first set of estimators is unwieldy to compute, while the second set is simple. The second set of estimators, though less efficient than the first, is more efficient than the usually employed estimators.

It is proved in subfield terminology that if $\mathcal{S}_1$ and $\mathcal{S}_2$ are two sufficient subfields and K is a set common to $\mathcal{S}_1$ and $\mathcal{S}_2$, then $\mathcal{S}_1 K + \mathcal{S}_2 K'$ is also a sufficient subfield. Hence the subfield $\mathcal{S}_1 K + \mathcal{S}_2 K'$ can be used to derive improved estimators by Rao-Blackwell theorem. Generalisation of this is also given in case of countable number of subfields. Application of this result to sampling with unequal probabilities is given.

1. INTRODUCTION

Consider a population containing N units. Let y_j be some real-valued characteristic of the j-th population unit in which we are interested.[1] Suppose that a sample of size n is drawn from the above population with unequal probabilities of selection (with replacement), P_j being the probability of selection associated with the j-th population unit ($\Sigma P_j = 1$). If for the i-th sample unit, we record its Y-characteristic y_i, probability of selection p_i and unit-index u_i, then the sample of observation is

$$S = (x_1, x_2, ..., x_n),$$

where $x_i = (y_i, p_i, u_i)$.[2]

It has been shown by Basu (1958) that the 'order-statistic'

$$T = (x_{(1)}, x_{(2)}, ..., x_{(\nu)})$$

(where $x_{(1)}, x_{(2)}, ... x_{(\nu)}$, are the distinct units in the sample arranged in ascending order of their unit-indices) is sufficient.

Therefore, if $g(S)$ is some estimator depending on the sample S, for any convex (downwards) loss function, an estimator uniformly better than $g(S)$ is given by $E[g(S)|T]$. In the subsequent sections this result is used to derive improved estimators of the population total and its square.

2. ESTIMATION OF THE POPULATION TOTAL

The usual estimator of the population total

$$Y = \Sigma Y_j$$

is given by

$$\bar{z} = \frac{1}{n} \Sigma z_i, \qquad \qquad ... \quad (2.1)$$

where

$$z_i = \frac{y_i}{p_i}.$$

---

[1] j varies from 1 to N, i from 1 to n and (i) from (1) to (ν) unless otherwise stated.

[2] Capital letters refer to the population and small letters to the sample.

Theorem 1 : *For any convex (downwards) loss function, an estimator uniformly better than $\bar{z}$ is given by*

$$\bar{z}_\nu = E[\bar{z}\,|\,T] = \Sigma\, C_{(i)} \frac{y_{(i)}}{p_{(i)}} \qquad \ldots \text{(2.2)}$$

where $\quad C_{(i)} = \dfrac{p_{(i)}[(p_{(1)}+\ldots+p_{(\nu)})^{n-1}-\Sigma_1^i(p_{(1)}+\ldots+p_{(r-1)})^{n-1}+\ldots(-)^{\nu-1}p_{(i)}^{n-1}]}{[(p_{(1)}+\ldots+p_{(\nu)})^n-\Sigma_1(p_{(1)}+\ldots+_{(\nu-1)})^n+\ldots(-)^{\nu-1}\Sigma_1 p_{(1)}^n]} \quad \ldots \text{(2.3)}$

the summations Σ_1 and Σ_1^i stand for all combinations of p's and all combinations of p's containing $p_{(i)}$ (chosen out of $p_{(1)}, p_{(2)}, \ldots, p_{(\nu)}$) respectively.

Proof : Obviously by Rao-Blackwell theorem, an estimator uniformly better than $\bar{z}$ is given by

$$E(\bar{z}\,|\,T) = E\left(\frac{y_1}{p_1}\,\Big|\,T\right) = \Sigma\, \frac{y_{(i)}}{p_{(i)}}\, P[x_1 = x_{(i)}\,|\,T]. \qquad \ldots \text{(2.4)}$$

But $\qquad P[x_1 = x_{(i)}\,|\,T] = \dfrac{p_{(i)}\Sigma'' \dfrac{(n-1)!}{\alpha_{(1)}!\ldots\alpha_{(\nu)}!}\, p_{(1)}^{\alpha_{(1)}} \ldots p_{(\nu)}^{\alpha_{(\nu)}}}{\Sigma' \dfrac{n!}{\alpha_{(1)}!\ldots\alpha_{(\nu)}!}\, p_{(1)}^{\alpha_{(1)}} \ldots p_{(\nu)}^{\alpha_{(\nu)}}} \qquad \ldots \text{(2.5)}$

where Σ' means summation over all integral α's such that

$$\alpha_{(i)} > 0 \text{ for } i = 1, \ldots, \nu \quad \text{and} \quad \alpha_{(1)}+\alpha_{(2)}+\ldots, +\alpha_{(\nu)} = n,$$

and Σ'' means summation over all integral α's such that

$$\alpha_{(i)} \geqslant 0,\ \alpha_{(i')} > 0 \quad \text{for} \quad i' \neq i = 1, 2, \ldots, \nu$$

and $\qquad\qquad \alpha_{(1)}+\alpha_{(2)}+\ldots,+\alpha_{(\nu)} = (n-1).$

It can be seen by induction over ν that

$$\Sigma' \frac{n!}{\alpha_{(1)}!\ldots\alpha_{(\nu)}!}\, p_{(1)}^{\alpha_{(\nu)}} \ldots p_{(\nu)}^{\alpha_{(\nu)}} = [(p_{(1)}+\ldots+p_{(\nu)})^n - \Sigma_1\,(p_{(1)}+\ldots+p_{(\nu-1)})^n$$
$$+\ldots(-)^{\nu-1}\,\Sigma_1\, p_{(1)}^n];$$

and

$$\Sigma'' \frac{(n-1)!}{\alpha_{(1)}!\ldots\alpha_{(\nu)}!}\, p_{(1)}^{\alpha_{(1)}} \ldots p_{(\nu)}^{\alpha_{(\nu)}} = [(p_{(1)}+\ldots+p_{(\nu)})^{n-1} - \Sigma_1^i(p_{(1)}+\ldots+p_{(\nu-1)})^{n-1}$$
$$+\ldots(-)^{\nu-1}\, p_{(i)}^{n-1}]. \qquad \ldots \text{(2.6)}$$

Using (2.4), (2.5) and (2.6), we get

$$\bar{z}_\nu = E(\bar{z}\,|\,T) = \Sigma\, C_{(i)} \frac{y_{(i)}}{p_{(i)}}$$

Hence the theorem is proved.

The above estimator, though better, is not very useful in large samples because of cumbersome computation of $C_{(i)}$'s. In Section 4, we shall derive a simpler estimator of Y better than $\bar{z}$. Table 1 gives exact expressions of z_ν for $n = 3$, 4 and 5. For $n = 1$ and 2, $\bar{z}_\nu$ and $\bar{z}$ are identical.

TABLE 1. $\bar{z}_\nu$ FOR $n=3$, 4 AND 5

$n \to$ ν	3	4	5
1	$\dfrac{y_{(1)}}{p_{(1)}}$	$\dfrac{y_{(1)}}{p_{(1)}}$	$\dfrac{y_{(1)}}{p_{(1)}}$
2	$\dfrac{\Sigma\, (2p_{(1)}+p_{(2)})\dfrac{y_{(1)}}{p_{(1)}}}{3(p_{(1)}+p_{(2)})}$	$\dfrac{\Sigma\left[(p_{(1)}+p_{(2)})^3-p_{(1)}^3\right]y_{(1)}}{\left[(p_{(1)}+p_{(2)})^4-p_{(1)}^4-p_{(2)}^4\right]}$	$\dfrac{\Sigma\left[(p_{(1)}+p_{(2)})^4-p_{(1)}^4\right]y_{(1)}}{\left[(p_{(1)}+p_{(2)})^5-p_{(1)}^5-p_{(2)}^5\right]}$
3	$\Sigma\,\dfrac{y_{(1)}}{p_{(1)}}$	$\dfrac{\Sigma\left[2p_{(1)}+p_{(2)}+p_{(3)}\right]\dfrac{y_{(1)}}{p_{(1)}}}{4[p_{(1)}+p_{(2)}+p_{(3)}]}$	$\dfrac{\Sigma\left[12p_{(1)}(p_{(1)}+p_{(2)}+p_{(3)})+4(p_{(2)}^2+p_{(3)}^2)+6p_{(2)}p_{(3)}\right]\dfrac{y_{(1)}}{p_{(1)}}}{5\left[4(p_{(1)}^2+p_{(2)}^2+p_{(3)}^2)+6(p_{(1)}p_{(2)}+p_{(2)}p_{(3)}+p_{(1)}p_{(3)})\right]}$
4	—	$\Sigma\,\dfrac{y_{(1)}}{p_{(1)}}$	$\dfrac{\Sigma\left[2p_{(1)}+p_{(2)}+p_{(3)}+p_{(4)}\right]\dfrac{y_{(1)}}{p_{(1)}}}{5\left[p_{(1)}+p_{(2)}+p_{(3)}+p_{(4)}\right]}$
5	—	—	$\Sigma\,\dfrac{y_{(1)}}{p_{(1)}}$

3. ESTIMATION OF Y^2

The problem of finding an unbiased estimator of Y^2 arises in most problems of variance estimation of estimators of Y. The usual estimator of Y^2 is

$$z_p = \frac{1}{n(n-1)} \sum_{i \neq i'=1}^{n} z_i\, z_{i'} \qquad \cdots \quad (3.1)$$

Theorem 2 : *For any convex (downwards) loss function, an estimator uniformly better than z_p is given by*

$$E(z_p|T) = \sum_{i=1}^{\nu} C_{(i,i)}\, z_{(i)}^2 + \sum_{i \neq i'=1}^{\nu} C_{(i,i')}\, z_{(i)}\, z_{(i')} \qquad \cdots \quad (3.2)$$

Remark : $\bar{z}_\nu$ when $\nu=(n-1)$ may be expressed in a simple form as $z_{(n-1)} = \dfrac{1}{n}\left\{\Sigma\dfrac{y_{(i)}}{p_{(i)}}+\dfrac{\Sigma\, y_{(i)}}{\Sigma\, p_{(i)}}\right\}$

317

where
$$C_{(i,i)} = \frac{p_{(i)}^2[(p_{(1)}+\ldots+p_{(\nu)})^{n-2}- \Sigma_1^i (p_{(1)}+\ldots+p_{(\nu-1)})^{n-2}+\ldots(-)^{\nu-1}p_{(i)}^{n-2}]}{[(p_{(1)}+\ldots+p_{(\nu)})^{n}- \Sigma_1(p_{(1)}+\ldots+p_{(\nu-1)})^{n}+\ldots (-)^{\nu-1}\Sigma_1 p_{(1)}^n]}$$

and $C_{(i,i')} =$

$$\frac{p_{(i)}p_{(i')}[(p_{(1)}+\ldots+p_{(\nu)})^{n-2}- \Sigma_1^{ii'}(p_{(1)}+\ldots+p_{(\nu-1)})^{n-2}+\ldots(-)^{\nu-2}(p_{(i)}+(p_{(i')})^{n-2}]}{[(p_{(1)}+\ldots+p_{(\nu)})^{n}- \Sigma_1(p_{(1)}+\ldots+p_{(\nu-1)})^{n}+\ldots (-)^{\nu-1}\Sigma_1 p_{(1)}^n]}$$

$$\ldots \quad (3.3)$$

the summations Σ_1 and Σ_1^i have been defined in (2.3) and the summation $\Sigma_1^{ii'}$ stands for all combinations of p's containing $p_{(i)}$ and $p_{(i')}$.

Proof :   Obviously

$$E(z_p|T) = E\left[\frac{1}{n(n-1)} \sum_{i\neq i'=1}^{n} z_i z_{i'} \Big| T \right] = E(z_1 z_2|T)$$

$$= \Sigma\ z_{(i)} z_{(i')} P[x_1 = x_{(i)}, x_2 = x_{(i')}|T]. \qquad \ldots \quad (3.4)$$

It is easy to see that

$$P[x_1 = x_{(i)}, x_2 = x_{(i)}|T] = \frac{p_{(i)}^2\ \Sigma'' \dfrac{(n-2)\,!}{\alpha_{(1)}\,!\,\ldots\,\alpha_{(\nu)}\,!}\ p_{(1)}^{\alpha_{(1)}}\ldots p_{(\nu)}^{\alpha_{(\nu)}}}{\Sigma'\ \dfrac{n\,!}{\alpha_{(1)}\,!\,\ldots\,\alpha_{(\nu)}\,!}\ p_{(1)}^{\alpha_{(1)}}\ldots p_{(\nu)}^{\alpha_{(\nu)}}} \qquad \ldots \quad (3.5)$$

and
$$P[x_1 = x_{(i)}, x_2 = x_{(i')}|T] = \frac{p_{(i)}p_{(i')}\Sigma''' \dfrac{(n-2)\,!}{\alpha_{(1)}\,!\,\ldots\,\alpha_{(\nu)}\,!}\ p_{(1)}^{\alpha_{(1)}}\ldots p_{(\nu)}^{\alpha_{(\nu)}}}{\Sigma'\ \dfrac{n\,!}{p_{(1)}\,!\,\ldots\,\alpha_{(\nu)}\,!}\ p_{(1)}^{\alpha_{(1)}}\ldots p_{(\nu)}^{\alpha_{(\nu)}}}, \qquad \ldots \quad (3.5.1)$$

where Σ', Σ'' have been defined in (2.5) and Σ''' means summation over all integral α's such that

$$\alpha_{(1)}+\ldots+\alpha_{(\nu)} = (n-2),\ \alpha_{(i)} \geqslant 0,\ \alpha_{(i')} \geqslant 0 \text{ and } \alpha_{(k)} > 0 \text{ for } k \neq i \neq i' = 1, \ldots, \nu.$$

It can be proved on lines similar to (2.6) (by induction over ν) that

$$\left.\begin{array}{l} P[x_{(1)} = x_{(i)}, x_{(2)} = x_{(i)} \mid T] = c_{(i,i)} \\[2mm] P[x_{(1)} = x_{(i)}, x_{(2)} = x_{(i')} \mid T] = c_{(i,i')} \end{array}\right\} \qquad \ldots \quad (3.6)$$

and

Using (3.4) and (3.6), we get

$$E[z_p \mid T] = \sum_{i=1}^{\nu} C_{(i,i)}\ z_{(i)}^2 + \sum_{i\neq i'=1}^{\nu} C_{(i,i)} z_{(i)} z_{(i')} \qquad \ldots \quad (3.7)$$

which proves the theorem.

Improved estimator of σ_z^2 :   The usual estimator of $\sigma_z^2 = \Sigma\ P_j \left(\dfrac{Y_j}{P_j} - Y \right)^2$ is

given by
$$s_z^2 = \frac{1}{(n-1)}\ \Sigma\ (z_i - \bar{z})^2 = \frac{1}{2n(n-1)}\ \sum_{i\neq i'=1}^{n}\ (z_i - z_{i'})^2.$$

Corollary 1 : *Thus an estimator uniformly better than s_z^2 is given by*

$$E[s_z^2 \mid T] = E\left[\frac{(z_1-z_2)^2}{2} \mid T\right] = \sum_{i \neq i'=1}^{\nu} C_{(i,i')}(z_{(i)}-z_{(i')})^2 \qquad \dots \quad (3.8)$$

Corollary 2 : *An unbiased estimator of $V(\bar{z}_\nu)$ is given by*

$$v(\bar{z}_\nu) = \bar{z}_\nu^2 - \sum_{i=1}^{\nu} C_{(i;i)} z_{(i)}^2 - \sum_{i \neq i'=1}^{\nu} C_{(i,i')} z_{(i)} z_{(i')}. \qquad \dots \quad (3.9)$$

Since this estimator is quite complicated for use in large samples, Basu (1958) has suggested the use of

$$\frac{1}{n(n-1)} \Sigma \ (z_i - \bar{z})^2 \qquad \dots \quad (3.10)$$

as an estimator of $V(\bar{z}_\nu)$. As it over-estimates $V(\bar{z}_\nu)$, we are always on the safe side to use (3.10) as our estimate.

The estimators derived in this and preceding sections, though superior to the usually employed estimators, are not of much use for large scale sample surveys owing to their cumbersome nature. In the next section, we give simpler estimators of Y and Y^2. These estimators, though less efficient than the above derived estimators, are superior to the usually employed estimators.

4. SIMPLE IMPROVED ESTIMATORS OF Y AND Y^2

Let us suppose that the observed samples are segregated into groups of equal p_i's. For instance, consider the problem of estimating the total yield of a crop from a sample of farms. Every sample-farm is to be selected with probability proportional to its area. Here, if some crude approximation (say correct to an acre) is used to measure their areas, we expect to get number of farms with same p_i in the sample. In the sequel, by the p-value of a unit, we mean the probability of selection associated with that unit. Let $p_{(1)}, p_{(2)}, \ldots, p_{(k)}$ be the distinct p-values of the sample units arranged in an increasing order of their magnitude. Let $n_{(i)}$ be the number of sample units having $p_{(i)}$ as their p-value. However, not all these $n_{(i)}$ units will be distinct, let $\nu_{(i)}$ be the number of distinct units among them. Now, if we arrange these $\nu_{(i)}$ distinct units in an increasing order of their unit-indices and call them $x_{(i1)}, x_{(i2)}, \ldots, x_{(i\nu_{(i)})}$, then it is not difficult to see that the statistic

$$T^* = [\{x_{(11)}, \ldots, x_{(1\nu_{(1)})}; n_{(1)}\}, \ldots, \{x_{(k1)}, \ldots, x_{(k\nu_{(k)})}; n_{(k)}\}] \qquad \dots \quad (4.1)$$

is sufficient.

It should be noted that if we take away the ancillary statistics $n_{(1)}, \ldots, n_{(k)}$ from the sufficient statistic T^*, then it reduces to the 'order-statistic' T defined in the earlier section. The 'unnecessarily wide' sufficient statistic T^* is used here for the purpose of deriving estimators of Y and Y^2 that are much simpler (though somewhat less efficient) than those considered in the previous section. Theorems 3 and 4 below give simple improved estimators of Y and Y^2 respectively.

Theorem 3 : *For any convex (downwards) loss function, an estimator uniformly better than $\bar{z}$ is given by*

$$\bar{z}_p^* = \frac{1}{n} \sum_{i=1}^{k} \frac{n_{(i)}}{p_{(i)}} \, \bar{y}_{\nu_{(i)}} \, , \qquad \ldots \ (4.2)$$

where

$$\bar{y}_{\nu_{(i)}} = \frac{1}{\nu_{(i)}} \sum_{r=1}^{\nu_{(i)}} y_{(ir)}.$$

Proof : Evidently an estimator uniformly better than $\bar{z}$ is given by

$$E(\bar{z} \,|\, T^*) = E\left(\frac{y_1}{p_1}\middle|\, T^*\right) \qquad \ldots \ (4.3)$$

Further, the probability of getting a sample with a given T^* is

$$P(T^*) = \frac{n!}{n_{(1)}! \ldots n_{(k)}!} \, p_{(1)}^{n_{(1)}} \ldots p_{(k)}^{n_{(k)}} C_{\nu_{(1)}}(n_{(1)}) \ldots C_{\nu_{(k)}}(n_{(k)}), \qquad \ldots \ (4.4)$$

where

$$C_{\nu_{(i)}}(n_{(i)}) = \nu_{(i)}^{n_{(i)}} - \binom{\nu_{(i)}}{1}(\nu_{(i)}-1)^{n_{(i)}} + \ldots (-)^{\nu_{(i)}-1}\binom{\nu_{(i)}}{\nu_{(i)}-1}1^{n_{(i)}};$$

$$(i = 1, \ldots, k)$$

and

$$P[x_1 = x_{(ir)} \,|\, T^*] = \frac{p_{(i)}\dfrac{(n-1)!}{n_{(1)}! \ldots (n_{(i)}-1)! \ldots n_{(k)}!} \, p_{(1)}^{n_{(1)}} \ldots p_{(i)}^{n_{(i)}-1} \ldots p_{(k)}^{n_{(k)}}}{\dfrac{n!}{n_{(1)}! \ldots n_{(i)}! \ldots n_{(r)}} \, p_{(1)}^{n_{(1)}} \ldots p_{(i)}^{n_{(i)}} \ldots p_{(k)}^{n_{(k)}}}$$

$$\times \frac{C_{\nu(1)}(n_{(1)}) \ldots \dfrac{C_{\nu(i)}(n_{(i)})}{\nu_{(i)}} \ldots C_{\nu(k)}(n_{(k)})}{C_{\nu(1)}(n_{(1)}) \ldots C_{\nu(i)}(n_{(i)}) \ldots C_{\nu(k)}(n_{(k)})}$$

$$= \frac{n_{(i)}}{n} \cdot \frac{1}{\nu_{(i)}} \, . \qquad \ldots \ (4.5)$$

From (4.3) and (4.5), it follows that

$$E(\bar{z} \,|\, T^*) = \frac{1}{n} \sum_{i=1}^{k} \frac{n_{(i)}}{p_{(i)}} \, \bar{y}_{\nu_{(i)}}$$

which completes the proof of the theorem.

A simple comparison of $\bar{z}_p^*$ and $\bar{z}$ will show that $\bar{z}_p^*$ will be superior to $\bar{z}$ if and only if the sample size is greater than two and at least three population units have the same p-value, otherwise z_p^* and $\bar{z}$ will be identical. It is not difficult to give a direct proof of the fact that $V(\bar{z}_p^*) \leqslant V(\bar{z})$. The strict sign of inequality holds only when the above condition is satisfied.

Theorem 4: *For any convex (downwards) loss function, an estimator uniformly better than*

$$z_p = \frac{1}{n(n-1)} \sum_{i \neq i'=1}^{n} \frac{y_i}{p_i} \cdot \frac{y_i'}{p_i'}$$

is given by

$$z_p^* = \frac{1}{n(n-1)} \left[\left\{ \left(\sum_{i=1}^{k} n_{(i)} \frac{\bar{y}_{\nu(i)}}{p_{(i)}} \right)^2 - \sum_{i=1}^{k} n_{(i)} \frac{\bar{y}_{\nu(i)}^2}{p_{(i)}^2} \right\} \right.$$
$$\left. - \sum n_{(i)}(n_{(i)}-1) \frac{C_{\nu(i)-1}(n_{(i)}-1)}{c_{\nu(i)}(n_{(i)})} \cdot \frac{s_{\nu(i)}^2}{p_{(i)}^2} \right], \quad \dots \quad (4.6)$$

where

$$s_{\nu(i)}^2 = \frac{1}{(\nu_{(i)}-1)} \sum_{r=1}^{\nu(i)} (y_{\nu(i)} - \bar{y}_{\nu(i)})^2, \quad and \quad C_{\nu(i)}(n_{(i)})$$

and $C_{\nu(i)-1}(n_{(i)}-1)$ *have meaning similar to those defined in (4.4).*

Proof: Obviously, an estimator uniformly better than z_p is given by

$$E[z_p \,|\, T^*] = E \left[\frac{y_1}{p_1} \cdot \frac{y_2}{p_2} \,\Big|\, T^* \right]. \quad \dots \quad (4.7)$$

Further, it can be shown that

$$P[x_1 = x_{(ir)}, \; x_2 = x_{(ir)} \,|\, T^*] = \frac{n_{(i)}(n_{(i)}-1)}{n(n-1)} \cdot \frac{C_{\nu(i)}(n_{(i)}-1)}{\nu_{(i)} C_{\nu(i)}(n_{(i)})} \,,$$

$$P[x_1 = x_{(ir)}, x_2 = x_{(ir')} \,|\, T^*] = \frac{n_{(i)}(n_{(i)}-1)}{n(n-1)} \times \left[\frac{C_{\nu(i)}(n_{(i)}) - C_{\nu(i)}(n_{(i)}-1)}{\nu_{(i)}(\nu_{(i)}-1) C_{\nu(i)}(n_{(i)}),} \right], \; (\nu \neq \nu')$$

and

$$P[x_1 = x_{(ir)}, \; x_2 = x_{(i'r')} \,|\, T^*] = \frac{n_{(i)} n_{(i')}}{n(n-1)} \cdot \frac{1}{\nu_{(i)}} \cdot \frac{1}{\nu_{(i')}} \; (i \neq i'). \quad \dots \quad (4.8)$$

Therefore,

$$E[z_p \,|\, T^*] = \sum_{i=1}^{k} \frac{n_{(i)}(n_{(i)}-1)}{n(n-1)} \sum_{r=1}^{\nu(i)} \frac{y_{i(r)}^2}{p_{(i)}^2} \frac{C_{\nu(i)}(n_{(i)}-1)}{\nu_{(i)} C_{\nu(i)}(n_{(i)})}$$
$$+ \sum_{i=1}^{k} \frac{n_{(i)}(n_{(i)}-1)}{n(n-1)} \sum_{r \neq r'=1}^{\nu(i)} \frac{y_{(ir)} y_{(ir')}}{p_{(i)}^2} \frac{[C_{\nu(i)}(n_{(i)}) - C_{\nu(i)}(n_{(i)}-1)]}{\nu_{(i)}(\nu_{(i)}-1) C_{\nu(i)}(n_{(i)})}$$
$$+ \sum_{i \neq i'=1}^{k} \frac{n_{(i)} n_{(i')}}{n(n-1)} \sum_{r=1}^{\nu(i)} \sum_{r'=1}^{\nu(i)} \frac{y_{(ir)}}{p_{(i)}} \frac{y_{(i'r')}}{p_{(i')}} \frac{1}{\nu_{(i)}} \frac{1}{\nu_{(i')}}. \quad \dots \quad (4.9)$$

Using the equality

$$\frac{C_{\nu(i)-1}(n_{(i)}-1)}{C_{\nu(i)}(n_{(i)})} s_{\nu(i)}^2 = \bar{y}_{\nu(i)}^2 - \frac{C_{\nu(i)}(n_{(i)}-1)}{\nu_{(i)} C_{\nu(i)}(n_{(i)})} \sum_{r=1}^{\nu(i)} y_{(ir)}^2$$
$$- \frac{[C_{\nu(i)}(n_{(i)}) - C_{\nu(i)}(n_{(i)}-1)]}{\nu_{(i)}(\nu_{(i)}-1) C_{\nu(i)}(n_{(i)})} \sum_{r \neq r'=1}^{\nu(i)} y_{(ir)} y_{(ir')},$$

and simplifying (4.9) we get

$$z_p^* = E[z_p | T^*] = \frac{1}{n(n-1)} \left[\left\{ \left(\sum_{i=1}^{k} n_{(i)} \frac{\bar{y}_{\nu_{(i)}}}{p_{(i)}} \right)^2 - \sum_{i=1}^{k} n_{(i)} \frac{\bar{y}_{\nu_{(i)}}^2}{p_{(i)}^2} \right\} \right.$$
$$\left. - \sum_{i=1}^{k} n_{(i)}(n_{(i)}-1) \cdot \frac{C_{\nu_{(i)}-1}(n_{(i)}-1)}{C_{\nu_{(i)}}(n_{(i)})} \cdot \frac{s_{\nu_{(i)}}^2}{p_{(i)}^2} \right].$$

This completes the proof.*

Corollary 3 : *It is easy to see that*

$$E[s_z^2 | T^*] = E\left[\frac{1}{(n-1)} \sum (z_i - \bar{z})^2 \mid T^* \right] = E\left[\frac{y_1^2}{p_1^2} \mid T^* \right] - E\left[\frac{y_1}{p_1} \cdot \frac{y_2}{p_2} \mid T^* \right]$$

$$= \sum_{i=1}^{k} \frac{n_{(i)}}{n} \cdot \frac{1}{\nu_{(i)}} \sum_{r=1}^{\nu_{(i)}} \frac{y_{(ir)}^2}{p_{(i)}^2} + \sum_{i=1}^{k} \frac{n_{(i)}(n_{(i)}-1)}{n(n-1)} \cdot \frac{C_{\nu_{(i)}-1}(n_{(i)}-1)}{C_{\nu_{(i)}}(n_{(i)})} \cdot \frac{s_{\nu_{(i)}}^2}{p_{(i)}^2}$$

$$- \frac{1}{n(n-1)} \left[\left(\sum_{i=1}^{k} n_{(i)} \frac{\bar{y}_{\nu_{(i)}}}{p_{(i)}} \right)^2 - \sum_{i=1}^{k} n_{(i)} \frac{\bar{y}_{\nu_{(i)}}^2}{p_{(i)}^2} \right] \qquad \dots \quad (4.10)$$

is a simple improved estimator of s_z^2.

Corollary 4 : *An unbiased estimator of $V(\bar{z}_\nu^*)$ is given by*

$$v(\bar{z}_\nu^*) = \bar{z}_\nu^{*2} + \sum_{i=1}^{k} \frac{n_{(i)}(n_{(i)}-1)}{n(n-1)} \cdot \frac{C_{\nu_{(i)}-1}(n_{(i)}-1)}{C_{\nu_{(i)}}(n_{(i)})} \cdot \frac{s_{\nu_{(i)}}^2}{p_{(i)}^2}$$

$$- \frac{1}{n(n-1)} \left[\left(\sum_{i=1}^{k} n_{(i)} \frac{\bar{y}_{\nu_{(i)}}}{p_{(i)}} \right)^2 - \sum_{i=1}^{k} n_{(i)} \frac{\bar{y}_{\nu_{(i)}}^2}{p_{(i)}^2} \right]. \qquad \dots \quad (4.11)$$

However, in practice it seems reasonable to use

$$\frac{1}{n(n-1)} \sum_i \left(\frac{z_i}{p_i} - \bar{z} \right)^2 \qquad \dots \quad (4.12)$$

as an estimator of $V(\bar{z}_\nu^*)$. First, because it is simple to compute, secondly, because it is always non-negative. Besides this, we are always on the safe side as it over-estimates the variance of $\bar{z}_\nu^*$.

---

* The estimator Z_p requires the computation of the ratio $\frac{C_{\nu-1}(n-1)}{C_\nu(n)}$. Values of $\frac{C_{\nu-1}(n-1)}{C_\nu(n)}$ can

be obtained from the relation $\frac{C_{\nu-1}(n-1)}{C_\nu(n)} = \frac{1}{\nu} - \frac{C_\nu(n-1)}{C_\nu(n)}$.

Values of $\frac{C_\nu(n-1)}{C_\nu(n)}$ have been tabulated for all ν and $n=1$ to 50 in a paper published elsewhere in the same issue (Pathak, 1962), pp. 287-302.

5. A REMARK ON RAO-BLACKWELL THEOREM

Let X be the sample space of all possible outcomes x. Let $\mathcal{S}$ be the field of subsets of X on which a set $P = \{p\}$ of probability measures is defined. Any statistic $T(x)$ [an $\mathcal{S}$-measurable function defined from X onto another space Y] generates a subfield $\mathcal{S}_T \subseteq \mathcal{S}$. The statistic T or the subfield $\mathcal{S}_T$ is called sufficient for P (Bahadur, 1954) if corresponding to each $\mathcal{S}$-measurable set A, there exists an $\mathcal{S}_T$-P-integrable function ϕ_A such that

$$P[x\epsilon A_0 \bigcap A] = \int_{A_0 \cap A} dp = \int_{A_0} \phi_A(x) dp \text{ for } A_0 \epsilon \mathcal{S}_T, \ p \epsilon P. \qquad \ldots \ (5.1)$$

It is known that any estimator $g(x)$ based on the sample x can be uniformly improved by taking the conditional expectation of $g(x)$ given a sufficient subfield. If several sufficient subfields are available, the minimum condensation of $g(x)$ is mostly obtained by employing the minimal sufficient subfield ($\mathcal{S}_1$ say). There are situations, e.g., in sample surveys where this minimal condensation is unwieldy and it is not possible for practical reasons to use this condensation. But it is sometimes possible to divide X into subsets $K(\epsilon \mathcal{S}_1)$ and K' such that the condensation is simple on K and unwieldy on K'. In such cases simpler condensation can be achieved with the help of some other subfield, $\mathcal{S}_2$, which contains the minimal one. It follows as a consequence of Theorem 5 (stated below) that condensation smaller than that of $\mathcal{S}_2$ (but larger than that of $\mathcal{S}_1$) can be obtained by employing $\mathcal{S}_1 K + \mathcal{S}_2 K'$ as the sufficient subfield, and this condensation will still have the merit of simplicity.

Theorem 5: *Let $\mathcal{S}_1$ and $\mathcal{S}_2$ be two sufficient subfields of $(X, \mathcal{S}, P)$, and K a set common to $\mathcal{S}_1$ and $\mathcal{S}_2$. Then the subfield**

$$\mathcal{S}_3 = \mathcal{S}_1 K + \mathcal{S}_2 K' \qquad \ldots \ (5.2)$$

is also sufficient.

Proof: Since $\mathcal{S}_1$ and $\mathcal{S}_2$ are two sufficient subfields, there exist for each $\mathcal{S}$-measurable set A, an $\mathcal{S}_1$-P-integrable function ϕ_{1A} and an $\mathcal{S}_2$-P-integrable function ϕ_{2A} such that

$$\int_{A_i \cap A} dp = \int_{A_i} \phi_{iA}(x) dp \quad \text{for } A_i \epsilon \mathcal{S}_i \quad (i = 1, 2,), p \epsilon P. \qquad \ldots \ (5.3)$$

Now for any $A_3 = A_1 K + A_2 K' \epsilon \mathcal{S}_3$, and for each $A \epsilon \mathcal{S}$

$$\int_{A_3 \cap A} dp = \int_{A_3} \phi_{3A} \, dp. \qquad \ldots \ (5.4)$$

where $\phi_{3A} = \phi_{1A} \chi_k + \phi_{2A} \chi k'$, and $\chi_k = 1 - \chi k'$ is the characteristic function of the set K.

It is easily seen that $\phi_{1A}(x) \chi_k$ and $\phi_{2A} \chi_k$ are both $\mathcal{S}_3$-P-integrable and thus ϕ_{3A} is $\mathcal{S}_3$-P-integrable. This with (5.4) implies that $\mathcal{S}_3$ is a sufficient subfield.

Corollary 5: *If $\mathcal{S}_1 \subseteq \mathcal{S}_2$, then $\mathcal{S}_1 \subseteq \mathcal{S}_3 \subseteq \mathcal{S}_2$.*

---

* $\mathcal{S}_3$ is the field of sets of the form $\mathcal{S}_3 = \{A_1 K + A_2 K'\}$, $A_i \epsilon \mathcal{S}_i$ $(i=1, 2)$.

Corollary 6 : Let $\mathcal{S}_1, \mathcal{S}_2, ..., \mathcal{S}_k$ be k sufficient subfields. If A_1, A_2, ..., A_k $[\bigcup_{i=1}^{k} A_i = X, A_i \cap A_j = \phi \,(i \neq j = 1, ..., k)]$ are k sets such that $A_i \in \mathcal{S}_i$ $(i = 1,$ ..., $k)$, then

$$S^* = A_1\mathcal{S}_1 + A_2\mathcal{S}_2 + ... + A_k\mathcal{S}_k \qquad \qquad .... \quad (5.5)$$

is also a sufficient subfield.

Corollary 7 : Let $g(x)$ be an estimator based on x. Let $\mathcal{S}_1$ and $\mathcal{S}_2$ be two sufficient subfields and K a set common to $\mathcal{S}_1$ and $\mathcal{S}_2$. Then for any convex (downwards) loss function, an estimator uniformly better than $g(x)$ is given by

$$E[g(x) | \mathcal{S}_3] = \begin{cases} E[g(x) | \mathcal{S}_1] & \text{if} \quad x \in K \\ E[g(x) | \mathcal{S}_2] & \text{otherwise,} \end{cases} \qquad ... \quad (5.6)$$

where $\mathcal{S}_3 = \mathcal{S}_1 K + {}_2 K'$.

This result is useful in estimation problems when the improved estimator $E[g(x) | \mathcal{S}_1]$ is difficult to compute for all $x \in X$. In such cases, this result may be utilized by employing subfield $\mathcal{S}_1$ and another subfield $\mathcal{S}_2$ such that $E[g(x) | \mathcal{S}_1]$ is simple to compute when $x \in K$ and $E[g(x) | \mathcal{S}_2]$ is simple to compute when $x \in K'$ $(K \in \mathcal{S}_1$ $i = 1, 2)$ and using (5.6) as the improved estimator. The resulting estimator will still be better than $g(x)$ and in addition will be simple to compute. Further if $\mathcal{S}_1 \subseteq \mathcal{S}_2$, this estimator will also be better than $E[g(x) | \mathcal{S}_2]$. For completeness a straightforward generalisation of Theorem 5 to the case of countable number of sufficient subfields is given below.

Theorem 6 : Let $\{\mathcal{S}_i\}$ be a countable number of sufficient subfields. Let $\{K_i\}$ be a sequence of mutually exclusive and exhaustive sets such that $K_i \in \mathcal{S}_i$ $(i = 1, 2, ...)$. Then

$$S^* = \sum_{i=1}^{\infty} K_i \mathcal{S}_i$$

is a sufficient subfield.

Proof : Since $\mathcal{S}_i$ is a sufficient subfield, there exists for each $A \in \mathcal{S}$ an $\mathcal{S}_i$-P-integrable function ϕ_{iA} such that

$$P[x \in A_i \cap A] = \int_{A_i \cap A} dp = \int_{A_i} \phi_{iA}(x) dp \text{ for } A_i \in \mathcal{S}_i \,(i = 1, ..., \infty), p \in P. \;\; ... \;\; (5.7)$$

Now for any $A^* = \Sigma A_i K_i \in S^*$ and for each $A \in \mathcal{S}$, we have from (5.7)

$$P[x \in A^* \cap A] = \lim_{n \to \infty} P[x \in (\sum_{i=1}^{n} A_i K_i) \cap A]$$

$$= \lim_{n \to \infty} \int_{\left(\sum_{i=1}^{n} A_i K_i\right)} \left[\sum_{i=1}^{n} \phi_{iA}(x) \chi_{K_i}(x) \right] dp$$

$$= \lim_{n \to \infty} \int_{A^*} \left[\sum_{i=1}^{n} \phi_{iA}(x) \chi_{K_i}(x) \right] dp. \qquad ... \quad (5.8)$$

Since $0 \leqslant f_n = \left(\sum\limits_{i=1}^{n} \phi_{iA}(x)\chi_{K_i}(x) \right) \leqslant 1$ a.e. (P), $\lim\limits_{n \to \infty} f_n = f_A^*$ exists a.e. (P). We, thus, have by Lebesgue's monotone convergence theorem

$$P[x \, \epsilon \, A^* \cap A] = \int_{A^*} f_A^*(x)dp. \qquad \qquad \ldots \text{ (5.9)}$$

Since for each n, f_n is S^*—P-integrable, $f_A^*(x)$ is S^*—P-integrable. Hence the theorem.

In the following section Theorem 5 is used in sampling with unequal probabilities to derive some simple improved estimators of Y.

6. APPLICATION TO SAMPLING WITH UNEQUAL PROBABILITIES

We have seen that $\bar{z}_\nu$ is uniformly better than $\bar{z}$. For $n = 3$, $\bar{z}_\nu$ can be expressed as

$$\bar{z}_\nu = \begin{cases} z_{(1)} & \text{if } \nu = 1, \\[2mm] \dfrac{1}{3}\left[z_{(1)} + z_{(2)} + \dfrac{y_{(1)} + y_{(2)}}{p_{(1)} + p_{(2)}} \right] & \text{if } \nu = 2 \\[2mm] \dfrac{1}{3}[z_{(1)} + z_{(2)} + z_{(3)}] & \text{if } \nu = 3. \end{cases} \qquad \ldots \text{ (6.1)}$$

It is not simple to compute $\bar{z}_\nu$ when $n > 3$ owing to cumbersome computation of $C_{(i)}$'s. However, if in a sample of size n, $\nu = (n-1)$, $\bar{z}_\nu$ is expressible in the simple form

$$\bar{z}_{(n-1)} = \frac{1}{n}\left[\sum\limits_{i=1}^{(n-1)} \frac{y_{(i)}}{p_{(i)}} + \frac{\sum\limits_{i=1}^{(n-1)} y_{(i)}}{\sum\limits_{i=1}^{(n-1)} p_{(i)}} \right]. \qquad \ldots \text{ (6.2)}$$

As a direct consequence of Corollary 7, it follows that a simple estimator uniformly better than $\bar{z}_\nu^*$ (and hence better than $\bar{z}$) is given by

$$\bar{z}_\nu' = \begin{cases} \bar{z}_\nu & \text{if } \nu = (n-1) \\[2mm] \bar{z}_\nu^* & \text{otherwise.} \end{cases} \qquad \ldots \text{ (6.3)}$$

Two points in favour of utilising z_ν' are: (i) it is as simple as $\bar{z}$ or $\bar{z}_\nu^*$ and (ii) it is more efficient than $\bar{z}_\nu^*$.

Another simple improved estimator of Y. Another simple improved estimator of Y can be derived by using the following sufficient statistic

$$T_2 = [(x_{(1)}, \alpha_{(1)}), \ldots, (x_{(\nu)}, \alpha_{(\nu)})] \qquad \ldots \text{ (6.4)}$$

where

$$\alpha_{(i)} = \begin{cases} \lambda_{(i)} & \text{if } \lambda_{(i)} > 2 \\[2mm] 1 & \text{otherwise} \quad (i = 1, \ldots; \nu) \end{cases}$$

and $\lambda_{(i)}$ is the number of times $x_{(i)}$ is included in the sample.

Assuming without any loss of generality that $\alpha_{(i)} = 1$ if $i = 1, \ldots, k$, and $\alpha_{(i)} > 1$ if $i = k+1, \ldots, \nu$, and $\sum\limits_{i=1}^{k} \lambda_{(i)} = m$, it can be shown that an estimator better than $\bar{z}$ is given by

$$\bar{z}_{\nu_{(2)}} = E[\bar{z} \mid T_2] = \frac{1}{n}\left[\sum\limits_{i=1}^{\nu} \alpha_{(i)}z_{(i)} + \sum\limits_{i=1}^{k} d_{(i)}z_{(i)} \right], \qquad \ldots \text{ (6.5)}$$

where $d_{(i)} = \dfrac{\Sigma_1^i\, p_{(1)} \cdots p_{(m-k)}}{\Sigma_1\, p_{(1)} \cdots p_{(m-k)}}$ the summations Σ_1 and Σ_1^i have been defined in (2.3)

and are taken over $p_{(1)} \ldots, p_{(k)}$.

In practical situations, it is much simpler to compute this estimator than to compute $\bar{z}_y$. For $m-k = 1$ and $m-k = 2$, $\bar{z}_{y_{(2)}}$ is given by

$$
\bar{z}_{y_{(2)}} =
\begin{cases}
\dfrac{1}{n}\left[\displaystyle\sum_{i=1}^{n} \alpha_{(i)} z_{(i)} + \dfrac{\sum_{i=1}^{k} y_{(i)}}{\sum_{i=1}^{k} p_{(i)}} \right] & \text{if } m-k = 1 \\[3ex]
\dfrac{1}{n}\left[\displaystyle\sum_{i=1}^{n} \alpha_{(i)} z_{(i)} + \dfrac{2k(y,\, p)}{k(p,\, p)} \right] & \text{if } m-k = 2
\end{cases}
\qquad \ldots \ (6.6)
$$

where $k(y,\, p) = \displaystyle\sum_{i=1}^{k} y_{(i)}\, p_{(i)} - (\Sigma y_{(i)})(\Sigma p_{(i)})$ and $k(p,\, p)$ is defined similarly.

In general when $(m-k)$ is large, this estimator may also involve some extra computation. If the statistician is not even in favour of this extra computation, the author, as a consequence of Corollary 7, recommends the following improved procedure of estimation

(i) use $\bar{z}$ if $(m-k)$ > 2 $\qquad \ldots \ (6.7)$
(ii) use $E[\bar{z} \mid T_2]$ if $(m-k) \leqslant 2$.

For estimating the variance of these estimators, author suggests (4.12) as an estimator.

7. Concluding remark

In case of large samples if one is interested in altogether dispensing with the extra computation, the observed sample of size n may be divided into sub-samples of sizes $n_1, n_2, \ldots, n_i$ etc., $(\Sigma n_i = n$ and $n_i = 3, 4$ or 5 etc.). This division should, however, be independent of sample observations. Each sub-sample may then be treated as a sample in itself and simple improved estimators may be obtained for each sub-sample by using the estimators given in the preceding section. The over-all improved estimator can now be obtained by averaging the estimators obtained from each sub-sample with weights proportional to the sub-sample sizes.

References

Bahadur, R. R. (1954) : Sufficiency and statistical decision function. *Ann. Math. Stat.*, **25**, 423-462.

Basu, D. (1958) : On sampling with and without replacement. *Sankhyā*. **20**, 287-294.

Fraser, D. A. S. (1957) : *Nonparametric Methods in Statistics*, John Wiley and Sons, New York.

Pathak, P. K. (1962) : On simple random sampling with replacement. *Sankhyā*, **24**, 287-302.

Paper received : August, 1961.

MISCELLANEOUS

AN EXISTENCE THEOREM IN SAMPLING THEORY*

By T. V. HANUMANTHA RAO

Indian Statistical Institute

SUMMARY. A (1,1) correspondence is established here, between sampling designs and sampling schemes, for sampling from a finite population. This result enables us to search for optimum sampling procedures in any particular case, through a unified general set-up.

1. INTRODUCTIONS

Let a finite population, of N units be given by

$$U_1, U_2, ..., U_i, ..., U_N. \qquad ... (1.1)$$

We give the following definitions.

Definition 1: A *sample* 's' from the above population is an ordered sequence

$$U_{i_1}, U_{i_2}, ..., U_{i_{n_s}}; \quad 1 \leqslant i_t \leqslant N, \quad \text{for} \quad 1 \leqslant t \leqslant n_s, \qquad ... (1.2)$$

where the i's need not necessarily be distinct and n_s is called the size of the sample 's'.

Definition 2: A *sampling scheme* is a process of selecting units one by one from the population (1.1) with pre-determined sets of probabilities of selection for individual units at each of the draws.

Definition 3: A *sample design* D is an arbitrary collection 'S' of samples 's' with an arbitrary probability measure P defined on it, according to which the samples should be drawn. We can write explicitly

$$D = D(S, P), \qquad ... (1.3)$$

where
$$\sum_{s \in S} P_s = 1.$$

It can be seen that this is the most general definition of a sample design.

It is known that any sampling scheme results in a unique sampling design, which is fully determined by it. That the converse also holds good, is shown in Section 2. We restrict our proof to the cases of practical importance when all the samples in the given design are of sizes $\leqslant m$, a fixed positive integer. However, the extension to the cases where this condition is not satisfied, such as sequential procedures, should not offer much difficulty.

A sample design is said to be completely specified if all possible samples including their permutations together with their respective probabilities of selection are specified. Examples of partially specified designs are (i) specification of all possible unordered samples

---

with their respective probabilities of selection and (ii) specification of the probabilities of inclusion in the sample of each of the units. Corresponding to any partially specified sample design, there may be many possibly different completely specified sampling designs.

2. Main results

Theorem : *There is a one to one correspondence between completely specified sample designs and sampling schemes.*

Proof : We introduce a new unit U_0 called the null unit, into our population. The occurrence of U_0 in a draw means that none of the units belonging to (1.1) is selected in that particular draw. We replace any sample

$$s = (U_{i_1}, U_{i_2}, ..., U_{i_{n_s}}) \quad \text{where} \quad n_s < m,$$

by

$$s' = (U_{i_1}, U_{i_2}, ..., U_{i_{n_s}}, U_0, U_0, ..., U_0),$$

such that the size of s' is equal to m. We attach the same probability to 's'' as the corresponding 's'. Let S' be the set of all s' and P' be the probability measure on S', as constructed above.

For any sampling scheme, let $\{p_{i_n}^{(n)} | (i_1, i_2, ..., i_{n-1})\}$, denote the probability of selecting U_{i_n} in the n-th draw, given that the first $(n-1)$ draws resulted in the selections of $U_{i_1}, U_{i_2}, ..., U_{i_{n-1}}$ successively. We shall now find p's such that the resulting sample design is the given design D. It would be sufficient to consider instead, the design $D' = (S', P')$. Then, let S'_{i_1} be the subset of S' consisting of all samples s' for which the first unit is U_{i_1}. Similarly, let $S'_{i_1 i_2}$ be the set of all samples which have U_{i_1} as their first unit and U_{i_2} as their second unit. The definitions of $S'_{i_1 i_2 i_3}$ etc. are now similar. In all the above definitions, the indices $i_1, i_2, ...,$ can be any integers not necessarily distinct, of the index set $0, 1, ..., N$. Let us define $p_{i_1}^{(1)}$ by

$$p_{i_1}^{(1)} = \sum_{s' \varepsilon S'_{i_1}} P_{s'}, \quad 1 \leqslant i_1 \leqslant N. \quad \quad ... (2.1)$$

Clearly, $\displaystyle\sum_{i_1=1}^{N} p_{i_1}^{(1)} = 1$, since $\displaystyle\bigcup_{i_1=1}^{N} S'_{i_1} = S'$ and $\displaystyle\sum_{s' \varepsilon S'} P_{s'} = 1.$

This defines the probabilities of selection of all units in the first draw. Let

$$\left\{ p_{i_2}^{(2)} | (i_1) \right\} = \begin{cases} \dfrac{\displaystyle\sum_{s' \varepsilon S'_{i_1 i_2}} P_{s'}}{p_{i_1}^{(1)}}, & \text{if} \quad p_{i_1}^{(1)} \neq 0 \quad \quad ... (2.2) \\[2em] 0 & \text{otherwise} \end{cases}$$

for $1 | \leqslant i_1 \leqslant N$, and $0 \leqslant i_2 \leqslant N.$

Since $\qquad S'_{i_1} = \bigcup_{i_2=0}^{N} S'_{i_1 i_2},$

it is clear that $\qquad \sum_{i_2=0}^{N} \left\{ p^{(2)}_{i_2} \,|\,(i_1) \right\} = 1 \qquad$ for $1 \leqslant i_1 \leqslant N,$

so that (2.2) completely defines the probabilities for selection in the second draw for all possible outcomes of the first draw. We observe here that the null unit cannot be the first unit of a sample s' since s' is the augmentation of a sample 's' of the design D.

Similarly, we define the probabilities of selection in the third draw by

$$\left\{ p^{(3)}_{i_3} \,|\,(i_1 i_2) \right\} = \begin{cases} \dfrac{\sum\limits_{s' \varepsilon S'_{i_1 i_2 i_3}} P_{s'}}{\sum\limits_{s' \varepsilon S'_{i_1 i_2}} P_{s'}} & \text{if } \sum\limits_{s' \varepsilon S'_{i_1 i_2}} P_{s'} \neq 0 \\[2em] 0 & \text{otherwise} \end{cases}$$

where the notation is clear. As before we have

$$\sum_{i_3=0}^{N} \left\{ p^{(3)}_{i_3} \,|\,(i_1, i_2) \right\} = 1 \qquad \text{for } 1 \leqslant i_1 \leqslant N$$
$$0 \leqslant i_2, i_3 \leqslant N.$$

This process can be continued until the m-th stage where it stops finally. This gives a well-defined sampling scheme giving rise to the design D' which is equivalent to the design D. That there is just one scheme giving rise to D' is clear because at each stage all the conditional probabilities $p^{(n)}_{i_n} \,|\, (i_1, i_2, \ldots, i_{n-1})$ should agree for both the schemes. This completes the proof of our assertion.

Remark 1 : The introduction of null unit in the population ensures that at each draw, we deal with a probability measure on the population

$$U_0, \, U_1, \, \ldots, \, U_N.$$

This removes undesirable ambiguity in some cases where we come across a draw in which there is a positive probability of no unit getting selected. For example, consider the population of 3 units

$$U_1, U_2 \text{ and } U_3.$$

Let S be the following set of samples

$(U_1);\ (U_2, U_1);\ (U_1, U_2);\ (U_3, U_2) \text{ and } (U_1, U_2, U_3),$ with probabilities 1/5 attached to each sample. Then

$$\sum_{i=1}^{3} \{ p_i^{(2)} \,|\,(i) \} = \frac{2}{5} \Big/ \frac{3}{5} = \frac{2}{3}$$

and with a probability $\frac{1}{5}$ we do not get any unit into the sample, which is an ambiguous situation. However, introducing the null unit U_0, the sample (U_1, U_0, U_0) carries a probability $\frac{1}{5}$ so that

$$\sum_{t=0}^{3} \{p_t^{(2)} | (1)\} = \frac{3}{5} \Big/ \frac{3}{5} = 1,$$

as it should be.

Remark 2 : We have seen above that for any general design there is a unique scheme which results in that design. We may call this scheme, "*the generating scheme of D*" and D the "*generated design*" of the scheme. However complicated a sampling design may be, we can always consider, conceptually at least, a scheme of drawing units one by one, which gives rise to the given design. Thus any sampling method which does not satisfy the definition of sampling scheme as given in Section 1, can be treated as equivalent to a suitably chosen sampling scheme.

BIOMETRIKA

Volume 49, Parts 1 and 2   **June, 1962**

CONTENTS

Miscellanea : Contributions by G. G. Alway ; H. E. Daniels ; W. J. Ewens and P. D. Finch ; J. Gani ; Gerald J. Glasser ; E. J. Hannan ; Shirley E. Hitchcock ; Joseph B. Keller ; A. M. Kshirsagar ; J. L. Mott ; R. S. Pinkham ; G. P. Steck and D. B. Owen ; P. S. Swamy ; J. S. Williams ; J. Wise.

Reviews
Other Books Received
Corrigenda

JOURNAL OF THE
ROYAL STATISTICAL SOCIETY

Series B (Methodological)

CONTENTS OF VOL. 24, PART 2, 1962

Obtainable through any bookseller or direct from the

ROYAL STATISTICAL SOCIETY

21 Bentinck Street, London, W. 1.

A NOTE ON MIXING PROCESSES

By K. R. PARTHASARATHY

Indian Statistical Institute

SUMMARY. In this paper it is shown that in the space of discrete stationary stochastic processes under the weak topology finite Markov Chains are dense and the set of weakly mixing processes is a dense G_δ.

The purpose of this note is to show that any real valued discrete stationary process can be approximated by means of strongly mixing Markov Chains and deduce that the set of weakly mixing processes is a set of the second category under the weak topology. This answers a question raised by Kolmogorov (1962).

DEFINITIONS AND NOTATIONS

Let R denote the real and R^I the countable product of R over all the integers. T denotes the shift transformation. $\mathcal{M}$ is the space of all distributions on R^I which are invariant under T. $\mathcal{M}$ is assigned the weak topology which makes it a complete separable metric space.

THEOREMS

Theorem 1 : *The set of strongly mixing Markov Chains is everywhere dense in $\mathcal{M}$.*

Proof : Consider points of the type x such that $T^k x = x$ for some k. The smallest k for which this is valid is called the period of x. The measure which assigns mass $1/k$ to the points $x, Tx, ..., T^{k-1}x$ is a periodic ergodic measure [as described by the author (Parthasarathy, 1961)]. Such measures have been proved to be dense in $\mathcal{M}$ (cf. Parthasarathy, 1961). From this we deduce the following.

Consider sequences of the following type

$$x = (... x_{-1}, x_0, x_1, x_2, ...)$$

where the numbers $x_0, x_1, ... x_{k-1}$ are all distinct and $x_r = x_{k+r}$ and measures with mass $1/k$ at $x, Tx, ..., T^{k-1}x$. Such measures are dense in $\mathcal{M}$.

We shall now approximate measures of this type by strongly mixing Markov Chains. To this end, we consider the Markov Chains with states $x_0, x_1, ..., x_{k-1}$, transition matrix

$$\begin{pmatrix} \epsilon/k-1 & 1-\epsilon & \epsilon/k-1 & ... & \epsilon/k-1 \\ \epsilon/k-1 & \epsilon/k-1 & 1-\epsilon & \epsilon/k-1 & ... & \epsilon/k-1 \\ ... & ... & ... & ... & ... & ... \\ 1-\epsilon & \epsilon/k-1 & \epsilon/k-1 & ... & \epsilon/k-1 & ... & \epsilon/k-1 \end{pmatrix}$$

and initial distribution $(1/k, 1/k, ..., 1/k)$ for $x_0, x_1, ..., x_{k-1}$. These are strongly mixing Markov Chains. As $\epsilon \to 0$ these Markov Chains converge weakly to the distribution with mass $1/k$ at the points $x, Tx, ..., T^{k-1}x$. This proves Theorem 1.

Corollary : *In particular the set of weakly mixing distributions is everywhere dense.*

Theorem 2 : *The set of weakly mixing distributions is a set of the second category in* $\mathfrak{M}$.

Proof : Let $\mathfrak{M}_w$ denote the set of weakly mixing distributions on R^I. It is enough to show that $\mathfrak{M}_w$ is a G_δ. We shall just indicate the proof since the arguments go exactly on the same lines as in the case of ergodic measures (Parthasarathy, 1961). We choose a metric in R^I such that the space of bounded uniformly continuous functions is separable under the uniform topology. Such a possibility is shown by Varadarajan (1958). A distribution μ is weakly mixing if and only if, for every (real valued) bounded uniformly continuous function f on R^I,

$$\lim_{n \to \infty} \frac{1}{n} \sum_1^n | \int f(T^r x) f(x) d\mu - (\int f \, d\mu)^2 | = 0.$$

This condition can be replaced by

$$\lim_{n \to \infty} \frac{1}{n} \sum_1^n | \int f(T^r x) f(x) d\mu - (\int f \, d\mu)^2 |^2 = 0. \qquad ... (1)$$

Since $\mu \times \mu$ is invariant under $T \times T$ and

$$\frac{1}{n} \sum_1^n | \int f(T^r x) f(x) \, d\mu - (\int f \, d\mu)^2 |^2$$

$$= \frac{1}{n} \sum_1^n \int \int (f(T^r x) - Ef)(f(T^r y) - Ef) \cdot (f(x) - Ef)(f(y) - Ef) d\mu \times d\mu \qquad ... (2)$$

the limit as $n \to \infty$ of the expression (2) exists for every stationary distribution. Therefore in condition (1) we can replace $\lim_{n \to \infty}$ by $\lim_{n \to \infty} \inf$. Hereafter the proof is exactly the same as in the case of ergodic distributions.

REFERENCES

Kolmogorov, A. N. (1962) : Lecture delivered at the Indian Statistical Institute. April 1962.

Parthasarathy, K. R. (1961) : On the category of ergodic measures. *Illinois. J. Math.,* 5, No. 4, 648-656.

Varadarajan, V. S. (1958) : Weak convergence of measures on separable metric spaces. *Sankhyā,* 19, 15-13.

Paper received : April, 1962.

SOME RESULTS ON UNBIASED ESTIMATION

By ROBERT V. HOGG

and

ALLEN T. CRAIG

University of Iowa

SUMMARY. This paper is concerned with the existence of (i) unbiased statistics for certain parameters, and (ii) efficient statistics for certain parameters, when the underlying probability density function has a given exponential form.

1. INTRODUCTION

The problem of point estimation of parameters of distributions of probability has long interested mathematical statisticians, and the literature contains the important and interesting theorems of many investigators. In this paper, we wish to call attention to some additional results, which, in so far as we are aware, have not yet been pointed out. A particularly interesting theorem is one that is concerned with the distribution of joint efficient estimates of several parameters.

Throughout this paper, probability density function is abbreviated p.d.f.; and density means density with respect to Lebesgue measure.

2. UNBIASED ESTIMATION IN REGULAR CASES

Let $X_1, X_2, ..., X_n$ denote the items of a random sample from a non-degenerate distribution which has a p.d.f. $f(x; \theta)$ that depends on the single parameter θ, $\theta \epsilon \Omega$, where Ω contains a non-degenerate interval. We take $f(x; \theta)$ to be of the exponential form

$$f(x; \theta) = \exp[\theta K(x) + S(x) + q(\theta)], \quad a_0 \leqslant x \leqslant b_0,$$

zero elsewhere, where a_0 and b_0, finite or infinite, do not depend upon θ, and $K'(x) = D_x K(x)$ is continuous and not identically equal to zero. It is well known that $Y = \overset{n}{\underset{1}{\Sigma}} K(X_i)$ is a sufficient statistic for θ and that Y has the p.d.f.

$$g(y; \theta) = \begin{cases} \exp[\theta y + T(y) + nq(\theta)], & a \leqslant y \leqslant b, \\ 0, & \text{elsewhere.} \end{cases}$$

Again, a and b, finite or infinite, do not depend upon θ. We then have the following theorem that gives, provided they exist, the unique unbiased statistics, based on Y, for θ^m, $m = 1, 2, 3, ...$. The conclusion of this theorem is the same as the conclusion of a theorem of Washio-Morimoto-Ikeda (1956). However, the hypotheses of the two theorems are different. We believe it is much easier to verify, for a specified $g(y; \theta)$, the hypotheses of the following theorem than it is to verify the hypotheses of the earlier theorem.

Theorem 1 : *Given that the preceding conditions on $f(x; \theta)$ and $g(y; \theta)$ are satisfied. Let the derivatives $D_y^k\{exp\ T(y)\}$, $k = 0, 1, ..., m$, exist (a.e.P). Let*

$$W_k = (-1)^k \frac{D_y^k\{\exp[T(Y)]\}}{\exp[T(Y)]}$$

and let $E[W_k]$ exist for $k = 1, 2, ..., m$. A necessary and sufficient condition that $W_1, W_2, ..., W_m$ be the unbiased statistics for $\theta, \theta^2, ..., \theta^m$ is that

$$[\{\exp(\theta y)\}\ D_y^k\{\exp[T(y)]\}]_{y=a} = [\{\exp(\theta y)D_y^k\{\exp[T(y)]\}]_{y=b}$$

for $k = 0, 1, ..., m-1$.

Proof of sufficiency : The proof is by mathematical induction. If $k = 1$, we have

$$E\left[-\frac{D_y\{\exp[T(Y)]\}}{\exp[T(Y)]}\right] = \int_a^b -D_y\{\exp[T(y)]\}\exp[\theta y + nq(\theta)]dy.$$

If we integrate by parts, the right member of this equation may be written

$$[-\exp\{T(y) + \theta y + nq(\theta)\}]_a^b + \theta\int_a^b \exp[\theta y + T(y) + nq(\theta)]dy.$$

The condition stated in the theorem implies that this last expression is equal to θ. For the general k, we have

$$E\left[(-1)^k\frac{D_y^k\{\exp[T(Y)]\}}{\exp[T(Y)]}\right] = \int_a^{\blacksquare}(-1)^kD_y^k\{\exp[T(y)]\}\exp[\theta y + nq(\theta)]dy.$$

The right member of this equation may be written

$$[(-1)^kD_y^{k-1}\{\exp[T(y)]\}\exp\{\theta y + nq(\theta)\}]_a^b$$
$$+\theta\int_a^b(-1)^{k-1}D_y^{k-1}\{\exp[T(y)]\}\exp[\theta y + nq(\theta)]dy. \qquad \ldots \quad (2.1)$$

The condition stated in the theorem requires that this is equal to

$$\theta E\left[(-1)^{k-1}\frac{D_y^{k-1}\{\exp[T(Y)]\}}{\exp[T(Y)]}\right].$$

We now invoke the induction hypothesis. Then the last expression may be written $\theta(\theta^{k-1}) = \theta^k$. Completeness (Lehmann and Scheffé, 1955) of the family $\{g(y; \theta); \theta\epsilon\Omega\}$ insures uniqueness. This completes the proof of the sufficiency of the condition.

Proof of necessity : If expression (2.1) is equal to θ^k and if the integral in this expression is equal to θ^{k-1}, for $k = 1, 2, \ldots, m$, we have that

$$\exp[\theta y]D_y^{k-1}\{\exp[T(y)]\}$$

has the same value at $y = a$ as it does at $y = b$. This establishes Theorem 1.

We note from Theorem 1, provided our conditions are satisfied with $m \geqslant 1$, that

$$E\left[(-1)\frac{D_y\{\exp[T(Y)]\}}{\exp[T(Y)]}\right] = E[-T'(Y)] = \theta.$$

Moreover, with $m \geqslant 2$,

$$\theta^2 = E\left[(-1)^2\frac{D_y^2\{\exp[T(Y)]\}}{\exp[T(Y)]}\right] = E[\{T'(Y)\}^2 + T''(Y)],$$

and accordingly the variance of $-T'(Y)$ is

$$\text{var}\{-T'(Y)\} = E[\{-T'(Y)\}^2] - \theta^2 = -E[T''(Y)].$$

In Section 3 we discuss when, and only when, this variance is equal to the Rao-Cramér lower bound.

It is easily seen, from the proof of Theorem 1, that if appropriate changes are made in the hypotheses and if $f(x; \theta)$ is of the form

$$\exp[p(\theta)K(x) + S(x) + q(\theta)], \quad a_0 \leqslant x \leqslant b_0,$$

then

$$E\left[(-1)^k\frac{D_y^k\{\exp[T(Y)]\}}{\exp[T(Y)]}\right] = [p(\theta)]^k.$$

Further, if we use anti-derivatives of $\exp[T(y)]$ instead of derivatives, we can, given hypotheses similar to those of Theorem 1, find unique unbiased statistics for $[p(\theta)]^{-k}$.

The following illustrative example points up the importance of the hypotheses of Theorem 1. Let X have the p.d.f. $f(x; \theta) = \exp[\theta x + \ln x(1-x) + q(\theta)]$, $0 \leqslant x \leqslant 1$, zero elsewhere, where $\theta \neq 0$ and

$$\exp[q(\theta)] = \frac{\theta^3}{\theta + 2 + (\theta - 2)\exp[\theta]}.$$

Consider a random sample of size one and let $Y = X$. Then $g(y; \theta) = f(y; \theta)$ and $T(y) = \ln y(1-y)$. It is easily verified that $E[-T'(Y)] = E[(2Y-1)/Y(Y-1)] = \theta$. To compute, in accordance with Theorem 1, the variance of $-T'(Y)$, we would need to be able to estimate θ^2. But the conditions of the theorem are satisfied only for $m = 1$ and we cannot estimate θ^2. As a matter of fact, in this instance, the variance of $-T'(Y)$ does not exist.

3. EFFICIENT STATISTICS : ONE PARAMETER CASE

In this section we shall investigate the conditions under which the unbiased statistic for θ, namely $-T'(Y)$, is efficient; that is, when the variance of $-T'(Y)$ attains the lower bound of the Rao-Cramér (1945, 1946a) inequality. Specifically, the following theorem will be proved.

Theorem 2 : Let $X_1, X_2, ..., X_n$ denote a random sample from a distribution that has a p.d.f. which satisfies the conditions of Theorem 1 with $m \geqslant 2$. Let $Y = \sum_1 K(X_i)$ and let $-T'(Y)$ denote the unbiased statistic for θ. A necessary and sufficient condition that the variance of $-T'(Y)$ be equal to the Rao-Cramér lower bound is that $K(X)$ have a non-degenerate normal distribution.

Proof of necessity : Assume first that $-T'(Y)$ is efficient. Since all regularity conditions (Cramér, 1946a) are satisfied, the variance of $-T'(Y)$ is given by

$$\frac{1}{E\left[\left(\frac{\partial \ln g}{\partial \theta}\right)^2\right]} = \frac{1}{E\left[-\frac{\partial^2 \ln g}{\partial \theta^2}\right]} = -\frac{1}{nq''(\theta)};$$

and also

$$\frac{\partial \ln g(y; \theta)}{\partial \theta} = k_1[-T'(y) - \theta],$$

where k_1 does not depend upon y. With $g(y; \theta)$ of the form under consideration, the latter condition becomes

$$y + nq'(\theta) = k_1[-T'(y) - \theta].$$

Since $k_1 \neq 0$, $T'(y)$ is the linear function $\{-y/k_1 - [nq'(\theta) + k_1\theta]/k_1\}$. However, $T'(y)$ is a function of y alone; thus both k_1 and $nq'(\theta) + k_1\theta$ are constants free of θ. Then, since $q(\theta)$ does not depend upon n, we have $q'(\theta) = -c\theta + d$ or

$$q(\theta) = -\frac{c}{2}\theta^2 + d\theta + e,$$

where c, d and e are constants free of θ. Now the characteristic function of the distribution of $K(X)$ is

$$E[\exp\{itK(X)\}] = \exp[q(\theta) - q(\theta + it)]$$
$$= \exp[(c\theta - d)(it) - ct^2/2].$$

Thus, if $-T'(Y)$ is an efficient statistic for θ, then $K(X)$, and hence $Y = \sum_1^n K(X_i)$, has a normal distribution.

Proof of sufficiency :   If $K(X)$, and hence $Y = \overset{n}{\underset{1}{\Sigma}} K(X_i)$, has a normal distribution, we know that

$$f(x; \theta) = \exp[\theta K(x) + S(x) - c\theta^2/2 + d\theta + e]$$
$$= \exp\{\theta[K(x) + d] + [S(x) + e] - c\theta^2/2\}.$$

Without loss of generality, we can replace $[K(x) + d]$ by $K(x)$ and $y = [\overset{n}{\underset{1}{\Sigma}} K(x_i) + nd]$ by $y = \overset{n}{\underset{1}{\Sigma}} K(x_i)$ to obtain

$$g(y; \theta) = \exp\{\theta y + T(y) - nc\theta^2/2\}.$$

Since $g(y; \theta)$ is a normal p.d.f., we can differentiate

$$\int_{-\infty}^{\infty} g(y; \theta)\, dy = 1 \qquad\qquad \dots \ (3.1)$$

with respect to θ under the integral.   We find that

$$E(Y - nc\theta) = 0,$$

or $Y/(nc)$ is the function of Y that is the unbiased statistic for θ.   From uniqueness, we know that $-T'(Y) = Y/(nc)$.   The variance of this statistic is found by differentiating (3.1) twice with respect to θ to obtain

$$E[(Y - nc\theta)^2 - nc] = 0 \quad \text{or} \quad \text{var}[Y/(nc)] = 1/(nc).$$

But this variance is equal to the Rao-Cramér lower bound $-1/[nq''(\theta)] = 1/(nc)$.   Thus $-T'(Y) = Y/(nc)$ is the efficient statistic for θ and Theorem 2 is established.

If in the preceding discussion θ is replaced by $p(\theta)$, we see that the variance of the unbiased statistic for $p(\theta)$ is equal to the Rao-Cramér lower bound

$$\frac{[p'(\theta)]^2}{E\left[\left(\dfrac{\partial \ln g}{\partial \theta}\right)^2\right]} = -\ \frac{1}{n\,\dfrac{d^2 q(\theta)}{d[p(\theta)]^2}}$$

if, and only if, $K(X)$ has a non-degenerate normal distribution.

4.   Efficient statistics : multiparameter case

Let $(X_{1j}, X_{2j}, \dots, X_{pj})$, $j = 1, 2, \dots, n$ denote a random sample from a non-degenerate p-variate distribution having p.d.f. of the form

$$\exp\left\{\overset{p}{\underset{1}{\Sigma}} \theta_i K_i(x_1, \dots, x_p) + S(x_1, \dots, x_p) + q(\theta_1, \dots, \theta_p)\right\},$$

where the domain of positive density does not depend upon $(\theta_1, \dots, \theta_p)\epsilon\ \Omega$, where Ω contains a non-degenerate p-dimensional interval.   Let $K_1, \dots, K_p$ be such that the sufficient statistics

$$Y_i = \overset{n}{\underset{j=1}{\Sigma}} K_i(X_{1j}, \dots, X_{pj}), \qquad i = 1, \dots, p,$$

have a non-degenerate p-variate distribution with p.d.f. $g(y_1, \dots, y_p; \theta_1, \dots, \theta_p)$ of the form

$$\exp\left\{\overset{p}{\underset{i=1}{\Sigma}} \theta_i y_i + T(y_1, \dots, y_p) + nq(\theta_1, \dots, \theta_p)\right\}.$$

Under these conditions we can prove the following theorem.

Theorem 3 : *If the conditions stated above are satisfied, a necessary and sufficient condition that the unbiased statistics for $\theta_1, \ldots, \theta_p$, based on $Y_1, \ldots, Y_p$, are joint efficient statistics is that $K_i(X_1, \ldots, X_p)$, $i = 1, \ldots, p$, have a non-degenerate p-variate normal distribution.*

Proof of necessity : If $U_1(Y_1, \ldots, Y_p), \ldots, U_p(Y_1, \ldots, Y_p)$ are the joint efficient unbiased statistics for $\theta_1, \ldots, \theta_p$, we have (Cramèr, 1946b) that

$$\frac{\partial \ln g}{\partial \theta_i} = k_{i1}(U_1-\theta_1)+\ldots+k_{ip}(U_p-\theta_p), \; i=1, \ldots, p,$$

where the k_{ij} do not depend upon $y_1, \ldots, y_p$. Under our conditions this becomes

$$y_i+n\,\frac{\partial q}{\partial \theta_i} = \sum_{j=1}^{p} k_{ij}(U_j-\theta_j), \; i=1, \ldots, p.$$

Obviously, the left members of these p equations are linearly independent; thus also are the right members. Consequently the matrix $K = (k_{ij})$ of the coefficients is non-singular. Accordingly if $y = (y_1, \ldots, y_p)'$, $U = (U_1, \ldots, U_p)'$, $\theta = (\theta_1, \ldots, \theta_p)'$, and

$$\frac{\partial q}{\partial \theta} = \left(\frac{\partial q}{\partial \theta_1}, \ldots, \frac{\partial q}{\partial \theta_p} \right)';$$

we have from the above equations that

$$U = K^{-1}y+K^{-1}\left(n\,\frac{\partial q}{\partial \theta}+K\theta \right).$$

Hence $U_1, \ldots, U_p$ are linear functions of $y_1, \ldots, y_p$. Since $U_1, \ldots, U_p$ are functions of $y_1, \ldots, y_p$ alone, then K^{-1}, and thus K, and $\left(n\frac{\partial q}{\partial \theta}+K\theta \right)$ are matrices whose elements are constants that are free of θ. That is,

$$\frac{\partial q}{\partial \theta} = -c\theta+d,$$

where $c = (c_{ij})$ and $d = (d_1, \ldots, d_p)'$ are matrices whose elements are constants that are free of θ. Here $c = c'$ since the second mixed partial derivative of q with respect to θ_i and θ_j does not depend upon the order of differentiation. Accordingly,

$$q = -\tfrac{1}{2}\theta'c\theta+\theta'd+e$$

where e is a constant. The characteristic function of $K' = (K_1, \ldots, K_p)$ is then

$$E[\exp(it'K)] = \exp\{it'(c\theta-d)-\tfrac{1}{2}t'ct\},$$

where $t' = (t_1, \ldots, t_p)$ is real. This completes the proof of the necessity of the condition.

Proof of sufficiency : Since $K_1, \ldots, K_p$ have a non-degenerate p-variate normal distribution, we can assume, possibly after trivial substitutions, that

$$q(\theta_1, \ldots, \theta_p) = -\tfrac{1}{2}\theta'A\theta$$

where $\theta' = (\theta_1, \ldots, \theta_p)$ and $A = (a_{ij})$ is a positive definite real symmetric matrix of order p whose elements do not depend upon θ. If we differentiate

$$\int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \exp\left[\sum_{1}^{p} \theta_i y_i+T(y_1, \ldots, y_p)+nq(\theta_1, \ldots, \theta_p) \right] dy_1 \ldots dy_p = 1 \quad \ldots \; (4.1)$$

with respect to θ_i, we obtain

$$E\left(Y_i+n\,\frac{\partial q}{\partial \theta_i} \right) = 0, \; i=1, 2, \ldots, p,$$

or

$$E(Y_i) = n(a_{i1}\theta_1+\ldots+a_{ip}\theta_p).$$

That is, if $Y' = (Y_1, ..., Y_p)$ then

$$E(Y) = n A \theta.$$

Consequently, if $U = (1/n) A^{-1} Y$, where $U' = (U_1, ..., U_p)$,

then

$$E(U) = (1/n) A^{-1} E(Y) = \theta.$$

That is, U provides the unbiased statistics, based on Y, for θ. If we consider the second partial derivatives of (4.1), we have

$$E\left[\left(Y_i + n \frac{\partial q}{\partial \theta_i}\right)\left(Y_j + n \frac{\partial q}{\partial \theta_j}\right) + n \frac{\partial^2 q}{\partial \theta_i \partial \theta_j}\right] = 0$$

or

$$E\left[\left(Y_i + n \frac{\partial q}{\partial \theta_i}\right)\left(Y_j + n \frac{\partial q}{\partial \theta_j}\right)\right] = n a_{ij}, \quad i, j = 1, ..., p.$$

Thus the matrix of variances and covariances of Y is nA. Then the matrix of variances and covariances of U is

$$\left(\frac{1}{n} A^{-1}\right)(nA)\left(\frac{1}{n} A^{-1}\right) = \frac{1}{n} A^{-1}.$$

The coefficients of the equation of the ellipsoid of greatest possible concentration (Cramér, 1946b) about θ are

$$b_{ij} = nE\left[-\frac{\partial^2 \ln g}{\partial \theta_i \partial \theta_j}\right] = -n \frac{\partial^2 q}{\partial \theta_i \partial \theta_j} = n a_{ij},$$

or $B = nA$. Since the inverse of the matrix of variances and covariances of U is equal to B, we see that $U_1, ..., U_p$ are the joint efficient statistics for $\theta_1, ..., \theta_p$. This completes the proof of the theorem.

An interesting observation is the following. Let us suppose that we are sampling from a k-variate distribution of the form assumed in Theorem 3. However, let k be less than p, where p is the number of parameters on which the distribution depends. If we follow the preceding arguments, we see that $K_1(X_1, ..., X_k), ..., K_p(X_1, ..., X_k)$ must have a p-variate non-degenerate normal distribution if joint efficient statistics for $\theta_1, ..., \theta_p$ are to exist. However, with $k < p$, this is impossible; so, in instances like this, joint efficient statistics do not exist. To be specific, suppose a one variable distribution, of the form given, depends upon two parameters. Then joint efficient statistics for those two parameters do not exist.

REFERENCES

CRAMÉR, HARALD (1946a) : *Mathematical Methods of Statistics*, Princeton University Press.

———— (1946b) : Contributions to the theory of statistical estimation. *Skandinavisk Aktuarietidskrift*, 29, 85-94.

LEHMANN, E. L. and SCHEFFÉ, H. (1955) : Completeness, similar regions, and unbiased estimation. *Sankhyā*, 15, 219-236.

RAO, C. RADHAKRISHNA (1945) : Information and accuracy attainable in the estimation of statistical parameters. *Bull. Cal. Math. Soc.*, 37, 81-91.

WASHIO, YASUTOSHI ; MORIMOTO, HARUKI ; and IKEDA, NOBUYUKI (1956) : Unbiased estimation based on sufficient statistics. *Bull. Math. Stat.*, 6, 69-93.

Paper received : April, 1961.

A MINIMAL SUFFICIENT STATISTIC FOR A GENERAL CLASS OF DESIGNS

By DAVID L. WEEKS
Oklahoma State University
and
FRANKLIN A. GRAYBILL
Colorado State University

SUMMARY. This paper derives a minimal sufficient statistic for a class of designs which includes the balanced incomplete block and the partially balanced incomplete block design with two associate classes as a subset. The derivation is given assuming a variance component model. (Eisenhart Model II).

1. INTRODUCTION

The usefulness of finding a set of minimal sufficient statistics is given when we consider a theorem proved by Rao-Blackwell, that a minimum variance unbiased estimate must be a function of a sufficient statistic for the family of densities under consideration. A minimal sufficient statistic (which always exists) for a family of densities is desirable in that when this statistic has been obtained, we have all the "information" contained in the sample about the indexing parameter condensed as far as possible.

After obtaining such a set, the completeness of the distribution of the minimal sufficient statistic should be determined. If this distribution is complete, then the problem of minimum variance unbiased estimation is solved. If not, then some function of the minimal sufficient statistic must be used. Unfortunately, if there are two or more unbiased estimates of a function of the parameter which are functions of a minimal sufficient statistic, the Rao-Blackwell Theorem does not tell us which has minimum variance.

This paper exhibits a minimal sufficient statistic for a rather general class of designs under the assumption of an Eisenhart Model II (Eisenhart, 1947). The class of designs includes as a subset the balanced incomplete block designs, and the partially-balanced incomplete block designs.

The results for the group-divisible partially balanced incomplete block designs with two associate classes are given as an example of the application of the general results. It is interesting to note that the quantities normally calculated in an analysis of variance do not include all the statistics in the minimal sufficient statistic for those designs.

2

2. Model and assumptions

We shall assume the matrix model

$$Y = X\gamma + e$$

where $X = (X_0, X_1, X_2)$; $\gamma' = (\mu, \beta', \tau')$ with the dimensions of the matrices and partitions as follows :

$Y(n \times 1)$; $X(n \times (b+t+1))$; $\gamma (b+t+1 \times 1)$; $e(n \times 1)$; $\mu(1 \times 1)$; $\beta(b \times 1)$;

$\tau(t \times 1)$; $X_0(n \times 1)$ where $X_{0j} = 1$ for $j = 1, 2, ..., n$; $X_1(n \times b)$

where $x_{1ij} = 0$ or 1, $i = 1, 2, ..., b$; $j = 1, 2, ..., n$; $X_2(n \times t)$

where $x_{2ij} = 0$ or 1, $i = 1, 2, ..., t$; $j = 1, 2, ..., n$.

The following distributional assumptions will be made:

(1) e is distributed as the multivariate normal, mean vector φ, covariance matrix $\sigma^2 I_n$;

(2) β is distributed as the multivariate normal, mean vector φ, covariance matrix $\sigma_1^2 I_b$;

(3) τ is distributed as the multivariate normal, mean vector φ, covariance matrix $\sigma_2^2 I_t$;

(4) μ is a scalar constant;

(5) $\operatorname{cov}(\beta, \tau) = \varphi$, $\operatorname{cov}(\beta, e) = \varphi$, $\operatorname{cov}(\tau, e) = \varphi$;

where φ represents the null matrix (vector) and I_p the $p \times p$ identity matrix. We shall assume in this paper that :

(1) rank $(X) = b + t - 1$

(2) $X_1' X_1 = k I_b$

(3) $X_2' X_2 = r I_t$.

Additional terms will be defined as follows :

(1) $N = X_2' X_1 (N$ is the incidence matrix of the design).

(2) $A = X_2' - k^{-1} X_1 N'$;

(3) j_q^p will denote a $(p \times q)$ *matrix* with every element unity;

(4) j_1 will denote $(s \times 1)$ vector with every element unity.

(5) We shall adopt the following convention : If B is a symmetric $b \times b$ matrix of rank c and P^* is an orthogonal matrix which diagonalizes B, we shall denote the $b \times b$ diagonal matrix $P^{*'} B P^*$ by D_B^*. A partition of P^* which diagonalizes B so that the c *non-zero* characteristic roots are on the diagonal will be denoted by $P'BP = D_B$. Non-starred diagonal matrices are non-singular with this convention and the characteristic vectors which yield this diagonal matrix will not be starred.

(6) $f = n - b - t + 1$.

3. Constructing an Orthogonal Transformation

In this section we shall construct an orthogonal matrix so that an orthogonal transformation can be made on the vector Y in order to facilitate the determination of a set of sufficient statistics for this model.

Since XX' is a symmetric $n \times n$ positive semi-definite matrix, there exists an orthogonal matrix Q^* such that $Q^{*'}XX'Q^* = D^*_{XX'}$ where $D^*_{XX'}$ is diagonal with $b+t-1$ positive elements on the main diagonal and $n-b-t+1$ elements equal to zero.

If we partition Q^* into (Q_1, Q_2^*) we have

$$Q^{*'}XX'Q^* = \begin{pmatrix} Q_1' \\ Q_2^{*'} \end{pmatrix} XX'(Q_1 Q_2^*) = \begin{pmatrix} Q_1' \\ Q_2^{*'} \end{pmatrix}(X_0X_0' + X_1X_1' + X_2X_2')(Q_1 Q_2^*)$$

$$= \begin{bmatrix} D_{XX'} & \varphi \\ \varphi & \varphi \end{bmatrix}$$

where $D_{XX'}$ is $(b+t-1) \times (b+t-1)$ and rank $(b+t-1)$. It can be shown that $Q_2^{*'}X_0 = \varphi$; $Q_2^{*'}X_1 = \varphi$; $Q_2^{*'}X_2 = \varphi$. Let Q_2^* be the last $n-b-t+1$ columns of a matrix P. For notation convenience let $Q_2^* = P_4$.

We shall now construct the first $b+t-1$ columns of P and show that the resulting matrix is orthogonal. Let R be a $(n \times b+t-1)$ matrix partitioned into (R_1, R_2, R_3) where R_1, R_2, R_3 are $(n \times 1)$, $(n \times b-1)$ and $(n \times t-1)$ matrices respectively.

Consider first the matrix $A'A = (X_2' - k^{-1}NX_1')(X_2 - k^{-1}X_1N')$. Since $A'A$ is a symmetric $t \times t$ matrix, there exists an orthogonal matrix P_3^* (say) such that

$$P_3^{*'}A'AP_3^* = D^*_{A'A}$$

where $D^*_{A'A}$ is diagonal with the characteristic roots of $A'A$ on the main diagonal. Since the design was assumed connected, the rank of $A'A$ is $t-1$. Hence one of the elements of the diagonal of $D^*_{A'A}$ is zero. Corresponding to the unique characteristic root, zero, there is a unique (normalized) vector ρ such that $\rho'A'A\rho = 0$. Since $j_t'A'Aj_1^t = 0$, $\rho = t^{-1/2}j_1^t$. Partition and construct P_3^* such that $P_3^* = (P_{30}^* P_3)$ where $P_{30}^* = \rho$. Hence $P_3'A'AP_3 = D_{A'A}$ where $D_{A'A}$ is $(t-1) \times (t-1)$ of full rank and all diagonal elements are positive.

Let us assume that there are s distinct positive roots of $A'A$ denoted by $d_1, d_2, ..., d_s$ of multiplicities $m_1, m_2, ..., m_s$, respectively. Thus

$$\sum_{i=1}^{s} m_i = t-1.$$

Partition the $(t-1)$ orthogonal vectors of P_3 such that $P_3 = (P_{31}, P_{32}, ..., P_{3s})$ where P_{3i} is $(t \times m_i)$. Thus

$$P_3'A'AP_3 = \begin{bmatrix} P_{31}' \\ P_{32}' \\ \vdots \\ P_{3s}' \end{bmatrix} A'A(P_{31}P_{32} ... P_{3s}) = \begin{bmatrix} d_1 I_{m_1} & \varphi & ... & \varphi \\ \varphi & d_2 I_{m_2} & ... & \varphi \\ \vdots & & & \\ \varphi & \varphi & & d_s I_{m_s} \end{bmatrix}$$

Since $(D_{A'A}^{-1/2} P_3'A')(AP_3 D_{A'A}^{-1/2}) = I_{t-1}$, $D_{A'A}^{-1/2} P_3'A'$ is a set of $t-1$ orthogonal vectors of dimension n. Let the R_3' portion of R (of P) be equal to $D_{A'A}^{-1/2} P_3' A'$; i.e., $R_3' = D_{A'A}^{-1/2} P_3' A'$.

Since $A'A = rI_t - k^{-1}NN'$ the $t \times t$ orthogonal matrix P_3^* also diagonalizes NN' since $NN' = rkI_t - kA'A$, then

$$P_3^{*'}NN'P_3^* = P_3^{*'}(rkI_t - kA'A)P_3^* = rkI_t - kD_{A'A}^*$$

Thus

$$P_3^{*'}NN'P_3^* = D_{NN'}^* = \begin{bmatrix} rk & \varphi \\ \varphi & rkI_{t-1} - kD_{AA'} \end{bmatrix}$$

with

$$P_3'NN'P_3 = rkI_{t-1} - kD_{A'A}.$$

Let us adopt the convention that the characteristic roots of $A'A$ which are equal to r, be equal to d_1. Thus the multiplicity of the zero characteristic roots of NN' will be m_1.

The characteristic root rk of NN' is unique and since $\rho'NN'\rho = rk$, ρ is the characteristic vector of NN' which corresponds to the root rk. Now

$$P_3^{*'}NN'P_3^* = \begin{bmatrix} P_{30}' \\ P_{31}'^* \\ \vdots \\ P_{3s}' \end{bmatrix} NN'(P_{30}P_{31}^* ... P_{3s}) = \begin{bmatrix} rk & \varphi & \varphi \\ \varphi & \varphi m_1 & \varphi \\ \varphi & \varphi & D_{NN'} \end{bmatrix}$$

where
$$D_{NN'} = \begin{bmatrix} k(r-d_2) & 0 & . & . & . & 0 \\ 0 & k(r-d_3) & & & & . \\ . & & & & & . \\ . & & & & & . \\ . & & & & & . \\ 0 & 0 & & . . . & & k(r-d_s) \end{bmatrix}$$

Let $k(r-d_i) = d_i^*$; $i = 2, ..., r$.

Now the non-zero characteristic roots of $N'N$ are identical to the non-zero characteristic roots of NN' and of the same multiplicity. Hence, letting P_2^* denote an orthogonal $b \times b$ matrix which diagonalizes $N'N$, we may write

$$P_2^{*\prime} N' N P_2^* = \begin{bmatrix} rk & \varphi & & \varphi \\ \varphi & \varphi_{m_1 + m_1'} & & \varphi \\ \varphi & \varphi & & D_{NN'} \end{bmatrix}$$

where $m_1' = b - t$ (note m_1' may be less than, greater than, or equal to zero). Partition P_2^* into $(P_{20} P_{21}^* P_2)$ where P_{20}, P_{21}^*, and P_2 are of dimension $b \times 1$, $(b \times (m_1 + m_1'))$ and $b \times \sum_{i=2}^{s} m_i$ respectively, where

$$P_2^* N' N P_2^* = \begin{bmatrix} P_{20}' \\ P_{21}^{*\prime} \\ P_2' \end{bmatrix} N'N (P_{20} P_{21}^* P_2) = \begin{bmatrix} rk & \varphi & & \varphi \\ \varphi & \varphi_{m_1 + m_1'} & & \varphi \\ \varphi & \varphi & & D_{NN'} \end{bmatrix}$$

The matrix $D_{NN'}^{-1/2} P_3' N$ is a set of $\sum_{i=2}^{s} m_i$ orthogonal vectors of dimension b and are the characteristic vectors which diagonalize $N'N$ to give the non-zero characteristic roots of $N'N$ excepting rk. Thus

$$D_{NN'}^{-1/2} P_3' N (N'N) N' P_3 D_{NN'}^{-1/2} = D_{NN'}$$

and the characteristic vectors corresponding to the non-zero characteristic roots of $N'N$ can thus be expressed in terms of the characteristic vectors of the non-zero characteristic roots of NN'.

Corresponding to the unique characteristic root rk of $N'N$, is the characteristic vector $b^{-1/2} j_1^b$ since $b^{-1/2} j_b' N'N j_1^b b^{-1/2} = rk$.

Now we are ready to define the orthogonal matrix P. Let

$$R_1' = n^{-1/2} j_n^1$$

$$R_2' = \begin{bmatrix} k^{-1/2} P_{21}^{*\prime} X_1' \\ k^{-1/2} P_2' X_1' \end{bmatrix} = \begin{bmatrix} k^{-1/2} P_{21}^{\prime*} X_1' \\ k^{-1/2} D_{NN'}^{-1/2} P_3' N X_1' \end{bmatrix}$$

Then

$$P' = \begin{bmatrix} n^{-1/2} j_n^1 \\[6pt] k^{-1/2} P_{21}^* X_1' \\[6pt] k^{-1/2} D_{NN'}^{-1/2} P_3' N X_1' \\[6pt] D_{A'A}^{-1/2} P_3' A' \\[6pt] P_4' \end{bmatrix}$$

It can be shown that the matrix P defined in this manner is orthogonal.

4. A SET OF SUFFICIENT STATISTICS FOR THE VARIANCE COMPONENTS

The $n \times 1$ vector Y is distributed as the multivariate normal with mean vector $X_0\mu = j_1^n\mu = \bar{\mu}$ (say) and covariance matrix $X_1 X_1' \sigma_1^2 + X_2 X_2' \sigma_2^2 + \sigma^2 I_n = \Sigma$ (say).

The joint density of the elements of the vector Y is then

$$g(Y;\theta) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp \{-1/2(Y-\bar{\mu})'\Sigma^{-1}(Y-\bar{\mu})\}$$

where $\theta' = (\mu, \sigma_1^2, \sigma_2^2, \sigma^2)$.

Consider now

$$g(Y;\theta) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp \{-1/2(Y-\bar{\mu})'PP'\Sigma^{-1}PP'(Y-\bar{\mu})\}$$

$$= \frac{1}{(2\pi)^{1/2} |\Sigma|^{1/2}} \exp \{-1/2(P'Y-P'\bar{\mu})'P'\Sigma^{-1}P(P'Y-P'\bar{\mu})\}$$

where P is the orthogonal matrix defined in the previous section.

The form of $P'\Sigma^{-1}P$ and $P'(Y-\bar{\mu})$ are needed in order to be able to define a set of sufficient statistics for this family of density functions.

Since $\boldsymbol{P}'\boldsymbol{\Sigma}^{-1}\boldsymbol{P} = (\boldsymbol{P}'\boldsymbol{\Sigma}\boldsymbol{P})^{-1}$, let us determine the form of $\boldsymbol{P}'\boldsymbol{\Sigma}\boldsymbol{P}$ and then invert the resulting matrix. From the manner in which P was defined we have

$$\boldsymbol{P}'\boldsymbol{\Sigma}\boldsymbol{P} =$$

$$
\begin{bmatrix}
\sigma^2+k\sigma_1^2+r\sigma_2^2 & \varphi & \varphi & \varphi & \varphi & \varphi \\
\varphi & (\sigma^2+k\sigma_1^2)\boldsymbol{I}_{m_1+m_1'} & \varphi & \varphi & \varphi & \varphi \\
\varphi & \varphi & \boldsymbol{I}_p(\sigma^2+k\sigma_1^2)+k^{-1}\boldsymbol{D}_{\mathbf{NN}'}\sigma_2^2 & \varphi & k^{-1/2}\boldsymbol{D}_{\mathbf{NN}'}^{1/2}\tilde{\boldsymbol{D}}_{\mathbf{AA}'}^{1/2}\sigma_2^2 & \varphi \\
\varphi & \varphi & \varphi & (\sigma^2+r\sigma_2^2)\boldsymbol{I}_{m_1} & \varphi & \varphi \\
\varphi & \varphi & k^{-1/2}\boldsymbol{D}_{\mathbf{NN}'}^{1/2}\tilde{\boldsymbol{D}}_{\mathbf{A}'\mathbf{A}}\sigma_2^2 & \varphi & \sigma^2\boldsymbol{I}_p+\tilde{\boldsymbol{D}}^{\mathbf{A}'\mathbf{A}}\sigma_2^2 & \varphi \\
\varphi & \varphi & \varphi & \varphi & \varphi & \sigma^2\boldsymbol{I}_f
\end{bmatrix}
$$

where
$$p = \sum_{i=2}^{s} m_i \text{ and } \tilde{\boldsymbol{D}}_{\mathbf{A}'\mathbf{A}} = \begin{bmatrix} d_2\boldsymbol{I}_{m_2} & & \varphi \\ & \cdot & \\ \varphi & & d_s\boldsymbol{I}_{m_s} \end{bmatrix}$$

More explicitly, we have :

(1) $(\sigma^2+k\sigma_1^2)\boldsymbol{I}_p+k^{-1}\boldsymbol{D}_{\mathbf{NN}'}\sigma_2^2 =$

$$
\begin{bmatrix}
(\sigma^2+k\sigma_1^2+k^{-1}d_2^\bullet\sigma_2^2)\boldsymbol{I}_{m_2} & \varphi & \cdots & \varphi \\
\varphi & (\sigma^2+k\sigma_1^2+k^{-1}d_3^\bullet\sigma_2^2)\boldsymbol{I}_{m_3} & \cdots & \varphi \\
\vdots & \vdots & \cdots & \vdots \\
\varphi & \varphi & \cdots & (\sigma^2+k\sigma_1^2+k^{-1}d_s^\bullet\sigma_2^2)\boldsymbol{I}_{m_s}
\end{bmatrix}
$$

(2) $\sigma^2\boldsymbol{I}_p+\boldsymbol{D}_{\mathbf{A}'\mathbf{A}}\sigma_2^2 =$

$$
\begin{bmatrix}
(\sigma^2+d_2\sigma_2^2)\boldsymbol{I}_{m_2} & \varphi & \vdots & \varphi \\
\varphi & (\sigma^2+d_3\sigma_2^2)\boldsymbol{I}_{m_3} & \vdots & \varphi \\
\vdots & \vdots & & \\
\varphi & \varphi & & (\sigma^2+d_s\sigma_2^2)\boldsymbol{I}_{m_s}
\end{bmatrix}
$$

(3) $k^{-1/2}\boldsymbol{D}_{\mathbf{NN}'}^{1/2}\tilde{\boldsymbol{D}}_{\mathbf{A}'\mathbf{A}}^{1/2}\sigma_2^2 =$

$$
\begin{bmatrix}
\sqrt{\dfrac{d_2^\bullet d_2}{k}}\,\boldsymbol{I}_{m_2}\,\sigma_2^2 & \varphi & \cdots & \varphi \\
\varphi & \sqrt{\dfrac{d_3^\bullet d_3}{k}}\,\boldsymbol{I}_{m_3}\,\sigma_2^2 & \cdots & \varphi \\
\vdots & \vdots & & \\
\varphi & \varphi & & \sqrt{\dfrac{d_s^\bullet d_s}{k}}\,\boldsymbol{I}_{m_s}\,\sigma_2^2
\end{bmatrix}
$$

345

The inverse of $(P' \Sigma P)$ is then seen to be

$$(P'\Sigma'P)^{-1} = P'\Sigma^{-1}P =$$

$$\begin{bmatrix} (\sigma^2+k\sigma_1^2+r\sigma_2^2)^{-1} & \varphi & \varphi & \varphi & \varphi & \varphi \\ \varphi & (\sigma^2+k\sigma_1^2)^{-1}\,I_{m_1+m_1'} & \varphi & \varphi & \varphi & \varphi \\ \varphi & \varphi & U_{33} & \varphi & U_{35} & \varphi \\ \varphi & \varphi & \varphi & (\sigma^2+r\sigma_2^2)I_{m_1} & \varphi & \varphi \\ \varphi & \varphi & U_{53} & \varphi & U_{55} & \varphi \\ \varphi & \varphi & \varphi & \varphi & \varphi & \sigma^2 I_f \end{bmatrix}$$

where
$$\Delta_i = \sigma^2+k\sigma^2\sigma_1^2+r\sigma^3\sigma_2^2+kd_i\sigma_1^2\sigma_2^2$$

$$U_{33} = \begin{bmatrix} (\sigma^2+d_2\sigma_2^2)\Delta_2^{-1}I_{m_2} & \varphi & \ldots & \varphi \\ \varphi & (\sigma^2+d_3\sigma_2^2)\Delta_3^{-1}\,I_{m_3} & & \varphi \\ \varphi & \varphi & \ddots & \varphi \\ \varphi & \varphi & & \varphi \\ \varphi & \varphi & \ldots & (\sigma^2+d_s\sigma_2^2)\Delta_s^{-1}\,I_{m_s} \end{bmatrix}$$

$$U_{55} =$$
$$\begin{bmatrix} [\sigma^2+k\sigma_1^2+k^{-1}(r-d_2)\sigma_2^2]\Delta_2^{-1}I_{m_2} & \varphi & \ldots & \varphi \\ \varphi & [\sigma^2+k\sigma_1^2+k^{-1}(r-d_3)\sigma_2^2]\Delta_3^{-1}\,I_{m_2} & \ldots & \varphi \\ \cdot & \cdot & \ldots & \cdot \\ \cdot & \cdot & \ldots & \cdot \\ \cdot & \cdot & \ldots & \cdot \\ \varphi & \varphi & [\sigma^2+k\sigma_1^2+k^{-1}(r-d_s)\sigma_2^2]\Delta_s^{-1}I_{m_s} \end{bmatrix}$$

$$U_{35} = U_{53}' = \begin{bmatrix} -\sqrt{\dfrac{d_2\overline{d_2^*}}{k}}\Delta_2^{-1}I_{m_2}\ \sigma_2^2 & \varphi & \ldots & \varphi \\ \varphi & -\sqrt{\dfrac{d_3\overline{d_3^*}}{k}}\Delta_3^{-1}I_{m_3}\ \sigma_2^2 & & \varphi \\ \cdot & & \ddots & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \varphi & \varphi\ \ldots & & -\sqrt{\dfrac{d_s\overline{d_s^*}}{k}}\Delta_s^{-1}I_{m_s}\ \sigma_2^2 \end{bmatrix}$$

We may now determine the form of $P'(Y-\bar{\mu})$.

$$P'(Y-\bar{\mu}) = \begin{vmatrix} n^{-1/2}j_n^1 \\ k^{-1/2}P_{21}^{\bullet}X_1' \\ k^{-1/2}P_{22}'X_1' \\ \cdot \\ \cdot \\ \cdot \\ k^{-1/2}P_{2s}' \\ r^{-1/2}P_{31}^{\bullet}{}'A' \\ d_2^{-1/2}P_{32}'A' \\ \cdot \\ \cdot \\ \cdot \\ d_s^{-1/2}P_{3s}'A' \\ P_4' \end{vmatrix} \quad (Y-X_0\mu) = \begin{vmatrix} n^{1/2}(\bar{y}-\mu) \\ k^{-1/2}P_{21}^{\bullet}{}'X_1'Y \\ k^{-1/2}P_{22}'X_1'Y \\ \cdot \\ \cdot \\ \cdot \\ k^{-1/2}P_{2s}'X_1'Y \\ r^{-1/2}P_{31}^{\bullet}{}'A'Y \\ d_2^{-1/2}P_{32}'A'Y \\ \cdot \\ \cdot \\ \cdot \\ d_s^{-1/2}P_{3s}'A'Y \\ P_4'Y \end{vmatrix}$$

Making the above substitutions for $P'\Sigma^{-1}P$ and $P'(Y-\bar{\mu})$ in (1)

we have

$$g(Y;\theta) = \frac{1}{(2\pi)^{n/2}}\,|\Sigma|^{-1/2}\,\exp^{-1/2q}$$

where

$$q = \frac{n(\bar{y}-\mu)^2}{\sigma^2+k\sigma_1^2+r\sigma_2^2} + \frac{Y'X_1P_{21}^{\bullet}P_{21}^{\bullet}{}'X_1'Y}{k(\sigma^2+k\sigma_1^2)} + \sum_{i=2}^{s}\frac{Y'X_1P_{2i}P_{2i}'X_1'Y}{k\Delta_i(\sigma^2+d_i\sigma_2^2)^{-1}}$$

$$+ \frac{Y'AP_{31}^{\bullet}P_{31}^{\bullet}{}'A'Y}{r(\sigma^2+r\sigma_2^2)} + \sum_{i=2}^{s}\frac{Y'AP_{3i}P_{3i}'A'Y}{d_i\Delta_i(\sigma^2+k\sigma_1^2+k^{-1}(r-d_i)\sigma_2^2)^{-1}}$$

$$+ \frac{Y'P_4P_4'Y}{\sigma^2} - 2\sum_{i=2}^{s}\frac{Y'X_1P_{2i}P_{3i}'A'Y\sigma_2^2}{\Delta_i}\sqrt{\frac{r-d_i}{k}}.$$

347

According to Neyman (1935) the $3s+1$ statistics defined as follows :

$$s_1 = \bar{y}$$
$$s_2 = k^{-1}Y'X_1P_{21}^{\bullet}P_{21}^{\bullet\prime}X_1'Y$$
$$s_3 = r^{-1}Y'AP_{31}^{\bullet}P_{31}^{\bullet\prime}A'Y$$
$$s_4 = Y'P_4P_4'Y$$
$$s_{5i} = k^{-1}Y'X_1P_{3i}P_{2i}'X_1'Y$$
$$s_{6i} = d_i^{-1}Y'AP_{3i}P_{3i}'A'Y$$
$$s_{7i} = \sqrt{\frac{r-d_i}{k}}\, Y'X_1P_{2i}P_{3i}'A'Y$$

$$\left.\right\} \quad i = 2, 3, ..., s$$

are a set of sufficient statistics for θ in this family of density functions.

Since $P_{2i} = (r-d_i)^{-1/2}P_{3i}'\, N$ we may alternatively write s_{5i} and s_{7i} as $[k(r-d_i)]^{-1}\, Y'X_1N'P_{3i}P_{3i}'NX_1'Y$ and $k^{-1}Y'X_1NP_{3i}P_{3i}'A'Y$, respectively.

4. Minimality of the set of sufficient statistics

We shall show in this section that the $3s+1$ statistics defined in Section 3 form a minimal set of sufficient statistics for $g(Y; \theta)$.

Lehmann and Scheffé (1950) have set forth a procedure by which a set of sufficient statistics may be shown to be minimal. In brief, their method involves proving that a function

$$K(Y, Y_0) = \frac{\mathscr{J}g(Y, \theta)}{\mathscr{J}g(Y_0; \theta)}$$

being independent of parameters implies $S = S_0$ where S is a vector-valued sufficient statistic for θ in $g(Y; \theta)$ and is usually considered to be a "Proposed" minimal set. S_0 is obtained from $g(Y_0; \theta)$ in the same way that S was obtained from $g(Y; \theta)$. $\mathscr{J}$ is an "operation" on a density function which reduces the dimension of the space of the sufficient statistics.

With these ideas in mind we proceed :

$$K(Y, Y_0) = \frac{\mathscr{J}g(Y; \theta)}{\mathscr{J}g(Y_0; \theta)} = \exp -\tfrac{1}{2}(q-q_0)$$

where we have defined $\mathscr{J}$ to be the orthogonal transformation P as defined in Section 2 and q as defined in Section 3.

Let

$$w_1 = n[(s_1-\mu)^2-(s_{10}-\mu)^2]$$
$$w_2 = s_2-s_{20}$$
$$w_3 = s_3-s_{30}$$
$$w_4 = s_4-s_{40}$$
$$w_{5i} = s_{5i}-s_{5i0} \qquad i = 2, ..., s$$
$$w_{6i} = s_{6i}-s_{6i0} \qquad i = 2, ..., s$$
$$w_{7i} = s_{7i}-s_{7i0} \qquad i = 2, ..., s$$

348

Then we may write $\exp-\frac{1}{2}(q-q_0)$ as

$$\exp-\frac{1}{2}\left(\sum_{i=1}^{4} f_i w_i + \sum_{j=5}^{7}\sum_{i=2}^{s} f_{ji}w_{ji}\right)$$

where

$$f_1 = (\sigma^2+k\sigma_1^2+r\sigma_2^2)^{-1}$$
$$f_2 = (\sigma^2+k\sigma_2^2)^{-1}$$
$$f_3 = (\sigma^2+r\sigma_2^2)^{-1}$$
$$f_4 = \sigma^{-2}$$
$$\left.\begin{aligned}f_{5i} &= (\sigma^2+d_i\sigma_2^2)\Delta_i^{-1}\\f_{6i} &= [\sigma^2+k\sigma_1^2+k^{-1}(r-d_i)\sigma_2^2]\Delta_i^{-1}\\f_{7i} &= -2\sigma_2^2\Delta_i^{-1}\end{aligned}\right\} \quad (i=2, 3, ..., s)$$

Now, $\exp-\frac{1}{2}(q-q_0)$ will be independent of parameters if the exponent in $K(Y, Y_0)$ is identically zero for every value of the parameters. To be zero the w_i and w_{ji} defined above must be identically zero. These can be identically zero if and only if the f_i's are a set of linearly independent functions.

In order to simplify the proof that these functions are linearly independent, consider the following transformation.

Let

$$x = \sigma^2; y = \sigma^2+k\sigma_1^2; z = \sigma^2+k\sigma_1^2+r\sigma_2^2.$$

Then

$$\Delta_i = xz+\frac{d_i}{r}(y-x)(z-y)$$

and

$$f_1 = \frac{1}{z}, f_2 = \frac{1}{y},$$

$$f_3 = \frac{1}{x+z-y}, f_4 = \frac{1}{x},$$

$$\left.\begin{aligned}f_{5i} &= \frac{x+\dfrac{d_i}{r}(y-z)}{xz+\dfrac{d_i}{r}(y-x)(z-y)}\\[2em]f_{6i} &= \frac{y+\left(\dfrac{r-d_i}{k}\right)(y-z)}{xz+\dfrac{d_i}{r}(y-x)(z-y)}\\[2em]f_{7i} &= \frac{\dfrac{1}{r}(y-z)}{xz+\dfrac{d_i}{r}(y-x)(z-y)}\end{aligned}\right\} \quad (i=2, 3, ..., s)$$

A common denominator of these $3s+1$ functions is

$$xyz(x+z-y)\prod_{i=2}^{s}\Delta_i = G \text{ (say)}.$$

Letting now
$$w = \frac{(y-x)(z-y)}{xz}; \quad \alpha_i = \frac{d_i}{r}; \quad \beta_i = \frac{(r-d_i)}{k};$$

$$\xi_h = \sum_{i_1 < i_2 < \ldots < i_h} \alpha_{i_1} \alpha_{i_2} \ldots \alpha_{i_h} \qquad\qquad i_q = 2, \ldots, s; \quad q = 1, 2, \ldots, h$$

$$\xi_h^\rho = \sum_{\substack{i_1 < i_2 < \ldots i_h \\ i_r \neq \rho}} \alpha_{i_1} \alpha_{i_2} \ldots \alpha_{i_h} \qquad\qquad i_q = 2, \ldots, s; \quad q = 1, 2, \ldots, h$$

we may write the $3s+1$ functions as follows :

$$Gf_1 = yx^s z^{s-1}(x+z-y) \sum_{h=0}^{s-1} w^h \xi_h$$

$$Gf_2 = x^s z^s (x+z-y) \sum_{h=0}^{s-1} w^h \xi_h$$

$$Gf_3 = yx^s z^s \sum_{h=0}^{s-1} w^h \xi_h$$

$$Gf_4 = yx^{s-1} z^s (x+y-z) \sum_{h=0}^{s-1} w^h \xi_h$$

$$Gf_{5i} = yx^{s-1} z^{s-1}(x+z-y)[x+\alpha_i(z-y)] \sum_{h=0}^{s-2} w^h \xi_h^i$$

$$= yx^{s-1} z^{s-1}(x+z-y)[x \sum_{h=0}^{s-2} w^h \xi_h^i + \alpha_i \sum_{h=0}^{s-2} (z-y)w^h \xi_h^i]$$

$$Gf_{6i} = yx^{s-1} z^{s-1}(x+z-y)[y+\beta_i(z-y)] \sum_{h=0}^{s-2} w^h \xi_h^i$$

$$= yx^{s-1} z^{s-1}(x+z-y)[y \sum_{h=0}^{s-2} w^h \xi_h^i + \sum_{h=0}^{s-2} \beta_i(z-y)w^h \xi_h^i]$$

$$Gf_{7i} = yx^{s-1} z^{s-1}(x+z-y)\frac{1}{r}(y-z) \sum_{h=0}^{s-2} w^h \xi_h^i.$$

To show that these $3s+1$ functions are linearly independent, we must show that

$$\sum_{i=1}^{4} f_i w_i + \sum_{j=5}^{7} \sum_{i=2}^{s} f_{ji} w_{ji}^* = 0$$

for every x, y, z, implies $w_i^* = 0, i = 1, 2, \ldots, 4; w_{ji}^* = 0, j = 5, 6, 7; \quad i = 2, 3, \ldots, s$.

We shall prove this is the case by equating coefficients of like terms on both sides of the identity above.

Since the terms $yx^{s+1}z^{s-1}w^{s-1}$, $xz^{s+1}w^{s-1}$, $yx^s z^s w^{s-1}$, and $y^2 x^{s-1}z^s w^{s-1}$ occur only in $f_1, f_2, f_3,$ and f_4, respectively, these four functions are mutually linearly independent and are linearly independent of $f_{5i}, f_{6i},$ and f_{7i} ($i = 2, 3, \ldots, s$). Hence we must now show that the $3(s-1)$ functions $f_{5i}, f_{6i},$ and f_{7i} ($i = 2, \ldots, s$) are mutually linearly independent.

Let

$$\Phi_k = xw^k$$
$$\psi_k = (z-y)w^k$$
$$\gamma_k = yw^k$$
$$\alpha_i \xi_h^i = \theta_h^i$$

and
$$\beta_i \xi_h^i = \delta_h^i.$$

Then

$$f_{5i} = yx^{s-1}z^{s-1}(x+z-y)\left[\sum_{h=0}^{s-2} \Phi_h \xi_h^i + \sum_{h=0}^{s-2} \psi_h \theta_h^i\right]$$

$$f_{6i} = yx^{s-1}z^{s-1}(x+z-y)\left[\sum_{h=0}^{s-2} \psi_h \delta_h^i + \sum_{h=0}^{s-2} \gamma_h \xi_h^i\right]$$

$$f_{7i} = yx^{s-1}z^{s-1}(x+z-y)\frac{1}{r}\sum_{h=0}^{s-1} \psi_h \xi_h^i$$

Choose the $3(s-1)$ terms $\Phi_0, \Phi_1, ..., \Phi_{s-2}, \psi_0, \psi_1, ..., \psi_{s-1}, \gamma_0, \gamma_1, ..., \gamma_{s-1}$. If we find the coefficients of these terms and equate these to what they are equal to in the identity and arrange them in matrix form we will have the system :

$$
\left[
\begin{array}{ccccccccccc}
\xi_0^1 & \xi_1^1 & \cdots & \xi_{s-2}^1 & \theta_0^1 & \theta_1^1 & \cdots & \theta_{s-2}^1 & 0 & 0 & \cdots & 0 \\
\xi_0^2 & \xi_1^2 & \cdots & \xi_{s-2}^2 & \theta_0^2 & \theta_1^2 & \cdots & \theta_{s-2}^2 & 0 & 0 & \cdots & 0 \\
\cdot & \cdot & & \cdot & \cdot & \cdot & & \cdot & \cdot & \cdot & & \cdot \\
\cdot & \cdot & & \cdot & \cdot & \cdot & & \cdot & \cdot & \cdot & & \cdot \\
\xi_0^{s-1} & \xi_1^{s-1} & \cdots & \xi_{s-2}^{s-1} & \theta_0^{s-1} & \theta_2^{s-1} & \cdots & \theta_{s-2}^{s-1} & 0 & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 & \delta_0^1 & \delta_1^1 & \cdots & \delta_{s-2}^1 & \xi_0^1 & \xi_1^1 & \cdots & \xi_{s-2}^1 \\
0 & 0 & \cdots & 0 & \delta_0^2 & \delta_1^2 & \cdots & \delta_{s-2}^2 & \xi_0^2 & \xi_1^2 & \cdots & \xi_{s-2}^2 \\
\cdot & \cdot & & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot & & \cdot \\
\cdot & \cdot & & \cdot & \cdot & \cdot & & \cdot & \cdot & \cdot & & \cdot \\
\cdot & \cdot & & \cdot & \cdot & \cdot & & \cdot & \cdot & \cdot & & \cdot \\
0 & 0 & \cdots & 0 & \delta_0^{s-1} & \delta_1^{s-1} & \cdots & \delta_{s-2}^{s-1} & \xi_0^{s-1} & \xi_1^{s-1} & & \xi_{s-2}^{s-1} \\
0 & 0 & \cdots & 0 & \frac{1}{r}\xi_0^1 & \frac{1}{r}\xi_1^1 & \cdots & \frac{1}{r}\xi_{s-2}^1 & 0 & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 & \frac{1}{r}\xi_0^2 & \frac{1}{r}\xi_1^2 & \cdots & \frac{1}{r}\xi_{s-2}^2 & 0 & 0 & \cdots & 0 \\
\cdot & \cdot & & \cdot & \cdot & \cdot & & \cdot & \cdot & \cdot & & \cdot \\
\cdot & \cdot & & \cdot & \cdot & \cdot & & \cdot & \cdot & \cdot & & \cdot \\
\cdot & \cdot & & \cdot & \cdot & \cdot & & \cdot & \cdot & \cdot & & \cdot \\
0 & 0 & \cdots & 0 & \frac{1}{r}\xi_0^{s-1} & \frac{1}{r}\xi_1^{s-1} & \cdots & \frac{1}{r}\xi_{s-2}^{s-1} & 0 & 0 & \cdots & 0
\end{array}
\right]
\left[
\begin{array}{c}
w_{51}^{*} \\
w_{52}^{*} \\
\cdot \\
\cdot \\
w_{5(s-1)}^{*} \\
w_{61}^{*} \\
w_{62}^{*} \\
\cdot \\
\cdot \\
\cdot \\
w_{6(s-1)}^{*} \\
w_{71}^{*} \\
w_{72}^{*} \\
\cdot \\
\cdot \\
\cdot \\
w_{7(s-1)}^{*}
\end{array}
\right]
=
\left[
\begin{array}{c}
0 \\
0 \\
\cdot \\
\cdot \\
0 \\
0 \\
0 \\
\cdot \\
\cdot \\
\cdot \\
0 \\
0 \\
0 \\
\cdot \\
\cdot \\
\cdot \\
0
\end{array}
\right]
$$

Let $H = \{\xi_j^i\}$, $K = \{\theta_j^i\}$, $M = \{\delta_j^i\}$, $W_k = \{w_{Ki}\}$. The above system can be written then as :

$$\begin{bmatrix} H & K & \varphi \\ \varphi & M & H \\ \varphi & \dfrac{1}{r}H & \varphi \end{bmatrix} \begin{bmatrix} W_5 \\ W_6 \\ W_7 \end{bmatrix} = \begin{bmatrix} \varphi \\ \varphi \\ \varphi \end{bmatrix}$$

If the determinant of the matrix on the left is non-zero, this implies the only solution to the above system is

$$\begin{bmatrix} W_5 \\ W_6 \\ W_7 \end{bmatrix} = \varphi.$$

It has been shown [Graybill and Hultquist (1961)], that $|H| \neq 0$. The absolute value of the determinant of the above system is then $\dfrac{1}{r}|H|^3 \neq 0$. Hence the only solution is

$$\begin{bmatrix} W_5 \\ W_6 \\ W_7 \end{bmatrix} = \begin{bmatrix} \varphi \\ \varphi \\ \varphi \end{bmatrix}$$

and the functions are linearly independent.

This then implies the minimality of the set of sufficient statistics set forth in Section 3.

5. An application to GD-PBIB's

In this section we shall apply the results of the previous sections to the group-divisible partially balanced incomplete block designs with two associate classes [Bose, Clatworthy and Shrikhande, (1953)].

The characteristic roots for the matrix $A'A$ for the singular, semi-regular, and regular GD-PBIB's with multiplicities of the roots are as follows.

	multiplicities		
	1	$m-1$	$m(n-1)$
singular	0	$\lambda_2 t k^{-1}$	r
semi-regular	0	r	$k^{-1}(rk-r+\lambda_1)$
regular	0	$\lambda_2 t k^{-1}$	$k^{-1}(rk-r+\lambda_1)$

Applying these results to the results of the previous sections, we have

singular $\qquad d_1 = r, \; m_1 = m(n-1); \; d_2 = \dfrac{\lambda_2 t}{k}, \; m_2 = m-1$

semi-regular $d_1 = r, \; m_1 = m-1; \; d_2 = \dfrac{1}{k}(rk-r+\lambda_1), \; m_2 = m(n-1)$

regular $\qquad d_2 = \dfrac{\lambda_2 t}{k}, \; m_2 = m-1; \; d_3 = \dfrac{1}{k}(rk-r+\lambda_1), \; m_3 = m(n-1).$

In the regular GD-PBIB, we have no roots of $A'A$ equal to r.

We then have the following sets of minimal sufficient statistics for the three designs considered here:

(A) singular

    (1) $\bar{y}$

    (2) $k^{-1} Y' X_1 P_{21}^* P_{21}^{*'} X_1 Y \qquad\qquad$ if $b > m$, and is not defined if $b = m$

    (3) $r^{-1} Y' A P_{31}^* P_{31}^{*'} A' Y$

    (4) $Y' P_4 P_4' Y$

    (5) $k^{-1} Y' X_1 P_{22} P_{22}' X_1 Y$

    (6) $\dfrac{k}{\lambda_2 t} \; Y' A P_{33} P_{32}' A' Y$

    (7) $\dfrac{(rk-\lambda_2 t)^{1/2}}{k} \; Y' X_1 P_{22} P_{32}' A' Y \quad$ or $\quad \dfrac{Y' X_1 N P_{32} P_{32}' A' Y}{k}$

(B) semi-regular

    (1) $\bar{y}$

    (2) $k^{-1} Y' X_1 P_{21}^* P_{21}^{*'} X_1' Y \qquad$ if $b > t-m+1$ and is not defined if $b = t-m+1.$

    (3) $r^{-1} Y' A P_{31}^* P_{31}^{*'} A' Y$

    (4) $Y' P_4 P_4' Y$

    (5) $k^{-1} Y' X_1 P_{22} P_{22}' X_1' Y$

    (6) $\dfrac{(rk-r+\lambda)}{k} \; Y' A P_{32} P_{32}' A Y$

    (7) $\dfrac{(r-\lambda_1)^{1/2}}{k} \; Y' X_1 P_{22} P_{32}' A' Y$ or $\dfrac{Y' X_1 N P_{32} P_{32}' A' Y}{k}$

353

(C) regular

(1) $\bar{y}$

(2) $k^{-1}Y'X_1P_{21}^{\bullet}P_{21}^{\bullet\bullet}X_1'Y$ if $b > t$ and is not defined if $b = t$

(3) $Y'P_4P_4'Y$

(4) $k^{-1}Y'X_1P_{22}P_{22}'X_1'Y$

(5) $k^{-1}Y'X_1P_{23}P_{23}'X_1Y$

(6) $\dfrac{k}{\lambda_2 t}Y'AP_{32}P_{32}'A'Y$

(7) $\dfrac{k}{rk-r-\lambda_1}Y'AP_{33}P_{33}'A'Y$

(8) $\dfrac{(rk-\lambda_2 t)^{1/2}}{k}Y'X_1P_{32}P_{32}'A'Y$ or $k^{-1}Y'X_1NP_{32}P_{32}'A'Y$

(9) $\dfrac{(r-\lambda_1)^{1/2}}{k}Y'X_1P_{23}P_{33}'A'Y$ or $k^{-1}Y'X_1NP_{33}P_{33}'A'Y$

6. Acknowledgement

References

Bose, R. C., Clatworthy, W. H., and Shrikhande, S. S. (1953): *Tables of Partially Balanced Designs with Two Associate Classes*, Report of North Carolina Agricultural Experiment Station, Number 107, Raleigh, North Carolina.

Eisenhart, C. (1947): The assumptions underlying the analysis of variance. *Biometrics*, **3**, 1-21.

Graybill, F. A. (1961): *Introduction to Linear Statistical Models*, Volume I, McGraw-Hill.

Graybill, F. A., and Hultquist, R. A. (1961): Theorems concerning Eisenhart's model II. *Ann. Math. Stat.*, **32**, 1, 261-269.

Lehmann, E. L. and Scheffé, H. (1950): Completeness, similar region, and unbiased estimation. *Sankhyā*, **10**, 305-340.

Neyman, J. (1935) : Sur un teorema concernente le cosidette statistiche sufficienti. *Giorn. Ist. Ital. Att.*, **6**, 320-334.

Paper received : April, 1962.

ON UNIQUENESS AND MAXIMA OF THE ROOTS OF LIKELI-HOOD EQUATIONS UNDER TRUNCATED AND CENSORED SAMPLING FROM NORMAL POPULATIONS

By S. A. D. C. DOSS

University of Poona, India

SUMMARY. It is shown that when estimating from truncated and censored samples both the parameters of normal populations jointly or only one of them the other being known, the maximum likelihood estimating equations possess a unique solution and that solution provides the absolute maximum of the likelihood function for all samples of any size.

1. INTRODUCTION

The problem of estimating the mean, μ, and the standard deviation, σ, of normal populations from truncated and censored samples has been considered by Cohen (1950, 1959), Gupta (1952), Hald (1949), and by many others. These authors have presented maximum likelihood estimating equations which are often solvable only by some iterative process. It has been tacitly assumed by them that their system of equations have a solution. The purpose of the present paper is to show that their system of equations possess a unique solution and that the solution maximises the likelihood function absolutely for all samples of any size. However, three different cases of the problem of estimation of the parameters of normal populations are considered, viz., (I) μ alone is estimated σ being known; (II) σ alone is estimated μ being known; and (III) both μ and σ being unknown they are estimated jointly. Throughout the paper by truncated and censored samples we mean doubly truncated and doubly censored samples.

Huzurbazar (1948) presented certain general necessary and sufficient condition for the existence of a solution of the likelihood equation for every sample of any size in the case of uniparametric distributions belonging to Pitman-Koopman's class. In Section 2, the condition given by Huzurbazar has been generalised to multiparametric distributions belonging to the same class. Since distributions belonging to Pitman-Koopman's class retain their property of possessing sufficient statistics even after truncation (see Pitman (1936) and Tukey (1949)), and since normal distribution is a member of this class, in the case of truncated samples to prove existence of a solution we only show that the condition given by Huzurbazar (1948) is satisfied in cases (I) and (II) and its generalisation is satisfied in case (III). When once these conditions are satisfied uniqueness and maxima of the solution easily follow from Huzurbazar's (1948, 1949) result that in the case of distributions admitting sufficient statistics the likelihood equations have a unique solution for every sample of any size and the solution does make the likelihood function a maximum. Incidentally, it may be remarked that in the light of the results of Huzurbazar (1949) and Tukey (1949), the results of Des Raj (1954) presented in his Section 5 need no proof.

We consider independent random observations from a normal population with probability density function (PDF)

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-(x-\mu)^2/2\sigma^2}, \quad -\infty < x < \infty.$$

If x_a and $x_\beta (x_\beta > x_a)$ are the points of truncation, we have a truncated sample when all record of observations below x_a and above x_β of a sample are omitted. On the other hand, we have a censored sample when the number of observations of a sample falling in the regions $(-\infty, x_a)$ and (x_β, ∞) are noted though their individual values are not measured. Throughout we shall adopt following notations :

$$t_i = \frac{x_i - \mu}{\sigma}, \quad Q_i = \frac{1}{\sqrt{2\pi}} e^{-t_i^2/2}, \quad \Phi_i = \int_{-\infty}^{t_i} Q \, dt.$$

2. Doubly truncated samples

Before showing that the necessary and sufficient condition given by Huzurbazar (1948) for the existence of a solution of the likelihood equation for all samples of any size is satisfied in cases (I) and (II), we shall generalise his condition for multiparametric distributions. In the uniparametric case the PDF of a distribution of the Pitman-Koopman's class is given as (see Pitman (1936))

$$f(x, \theta) = \exp \{u_1(\theta)v_1(x) + A(x) + B(\theta)\},$$

where u_1 and B are functions of θ, and v_1 and A are functions of x. In this case Huzurbazar's necessary and sufficient condition for the existence of a solution of the likelihood equation for every sample of any size is that the functions $v_1(x)$ and $-\dfrac{\partial B(\theta)}{\partial \theta} \Big/ \dfrac{\partial u_1(\theta)}{\partial \theta}$ should have the same range of values. In the multiparametric case the PDF of a distribution of the Pitman-Koopman's class is of the form (see Koopman (1936))

$$f(x, \theta_j) = \exp \left\{ \sum_{k=1}^{p} u_k(\theta_j)v_k(x) + A(x) + B(\theta_j) \right\},$$

where θ_j stands for $(\theta_1, \theta_2, ..., \theta_p)$ the parameters of the distribution. Now, the necessary and sufficient condition for the system of likelihood equations to have a solution for all samples of any size can be easily seen to be that the two transformations defined by

$$
\begin{bmatrix} v_1(x) \\ v_2(x) \\ \cdot \\ \cdot \\ \cdot \\ v_p(x) \end{bmatrix}
\quad \text{and} \quad -
\begin{bmatrix} \dfrac{\partial u_1(\theta_j)}{\partial \theta_1} & \dfrac{\partial u_2(\theta_j)}{\partial \theta_1} & \cdots & \dfrac{\partial u_p(\theta_j)}{\partial \theta_1} \\ \dfrac{\partial u_1(\theta_j)}{\partial \theta_2} & \dfrac{\partial u_2(\theta_j)}{\partial \theta_2} & \cdots & \dfrac{\partial u_p(\theta_j)}{\partial \theta_2} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \dfrac{\partial u_1(\theta_j)}{\partial \theta_p} & \dfrac{\partial u_2(\theta_j)}{\partial \theta_p} & \cdots & \dfrac{\partial u_p(\theta_j)}{\partial \theta_p} \end{bmatrix}^{-1}
\begin{bmatrix} \dfrac{\partial B(\theta_j)}{\partial \theta_1} \\ \dfrac{\partial B(\theta_j)}{\partial \theta_2} \\ \cdot \\ \cdot \\ \cdot \\ \dfrac{\partial B(\theta_j)}{\partial \theta_p} \end{bmatrix}
$$

should have the same range. We shall, now, show that Huzurbazar's condition is satisfied in cases (I) and (II), and its generalisation in case (III). We exclude the trivial cases where all the sample observations are equal to either x_a or x_β.

Case (I) : Estimation of μ alone σ being known. Without loss of generality let us take σ to be unity. Now the PDF of the population from which the samples are to be drawn is

$$f'(x, \mu) = \frac{1}{\sqrt{2\pi}(\Phi_\beta - \Phi_a)} e^{-(x-\mu)^2/2}, \ x_a \leqslant x \leqslant x_\beta. \qquad \cdots \ (2.1)$$

Let $x_1, x_2, \ldots, x_n$ be a sample of size n drawn from the above population. The likelihood function of the sample is

$$L_T(\mu) = \left(\frac{1}{\sqrt{2\pi}(\Phi_\beta - \Phi_a)}\right)^n \exp\left\{-\frac{1}{2}\sum_{i=1}^n (x_i - \mu)^2\right\}. \qquad \ldots \quad (2.2)$$

In this case
$$u_1(\mu) = \mu,$$

$$v_1(x) = x,$$

$$B(\mu) = -\log(\Phi_\beta - \Phi_a) - \frac{\mu^2}{2}, \qquad \ldots \quad (2.3)$$

and
$$-\frac{\partial B(\mu)}{\partial \mu} \Big/ \frac{\partial u_1(\mu)}{\partial \mu} = \mu - \frac{Q_\beta - Q_a}{\Phi_\beta - \Phi_a}.$$

By Cauchy's mean value theorem we have

$$\frac{Q_\beta - Q_a}{\Phi_\beta - \Phi_a} = -t' \qquad \ldots \quad (2.4)$$

where
$$t_a < t' < t_\beta.$$

Hence
$$-\frac{\partial B(\mu)}{\partial \mu} \Big/ \frac{\partial u_1(\mu)}{\partial \mu} = \mu + t' = x' \qquad \ldots \quad (2.5)$$

where $x_a < x' < x_\beta$. Thus we see that both $v_1(x)$ and $-\dfrac{\partial B(\mu)}{\partial \mu}\Big/\dfrac{\partial u_1(\mu)}{\partial \mu}$ have the same range

(x_a, x_β) and hence the likelihood equation $\dfrac{\partial \log L_T(\mu)}{\partial \mu} = 0$ has a unique solution that

maximises the likelihood function $L_T(\mu)$ for all samples of any size.

Case (II): *Estimation of σ alone μ being known.* In this case, without loss of generality we shall take μ to be zero. Then the PDF of the population from which, now, the samples are to be drawn is

$$f'(x, \sigma) = \frac{1}{\sigma\sqrt{2\pi}(\Phi_\beta - \Phi_a)} \, e^{-x^2/2\sigma^2}, \, x_a < x < x_\beta. \qquad \ldots \quad (2.6)$$

Let $x_1, x_2, \ldots, x_n$ be a sample of size n drawn from a normal population with PDF (2.6). The likelihood function of the sample is

$$L_T(\sigma) = \left(\frac{1}{\sigma\sqrt{2\pi}(\Phi_\beta - \Phi_a)}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n x_i^2\right\}. \qquad \ldots \quad (2.7)$$

This time,
$$u_1(\sigma) = -\frac{1}{2\sigma^2},$$

$$v_1(x) = x^2, \qquad\qquad\qquad \ldots \quad (2.8)$$

$$B(\sigma) = -\{\log \sigma + \log(\Phi_\beta - \Phi_a)\},$$

and
$$-\frac{\partial B(\sigma)}{\partial \sigma} \Big/ \frac{\partial u_1(\sigma)}{\partial \sigma} = \sigma^2\left\{1 - \frac{t_\beta Q_\beta - t_a Q_a}{\Phi_\beta - \Phi_a}\right\}.$$

Again by Cauchy's mean value theorem we have

$$-\frac{t_\beta Q_\beta - t_\alpha Q_\alpha}{\Phi_\beta - \Phi_\alpha} = 1 - t'^2. \qquad \ldots \text{(2.9)}$$

Hence,

$$-\frac{\partial B(\sigma)}{\partial \sigma} \bigg/ \frac{\partial u_1(\sigma)}{\partial \sigma} = \sigma^2 t'^2 = x'^2. \qquad \ldots \text{(2.10)}$$

Thus we see that $v_1(x)$ and $-\dfrac{\partial B(\sigma)}{\partial \sigma}\bigg/\dfrac{\partial u_1(\sigma)}{\partial \sigma}$ have the same range, and hence in this case also the likelihood equation $\dfrac{\partial \log L_T(\sigma)}{\partial \sigma} = 0$ possesses a unique solution and this solution provides an absolute maximum of $L_T(\sigma)$ for all samples of any size.

Case (III) : Joint estimation of μ and σ. In this case, the PDF of the population is

$$f'(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}(\Phi_\beta - \Phi_\alpha)} \, e^{-(x-\mu)^2/2\sigma^2}, \quad x_\alpha < x < x_\beta, \qquad \ldots \text{(2. 11)}$$

and the likelihood function of the sample is

$$L_T(\mu, \sigma) = \left(\frac{1}{\sigma\sqrt{2\pi}(\Phi_\beta - \Phi_\alpha)}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right\}. \qquad \ldots \text{(2.12)}$$

Now,

$$u_1(\mu, \sigma) = -\frac{1}{2\sigma^2},$$

$$u_2(\mu, \sigma) = \mu/\sigma^2,$$

$$\begin{bmatrix} v_1(x) \\ v_2(x) \end{bmatrix} = \begin{bmatrix} x^2 \\ x \end{bmatrix},$$

$$B(\mu, \sigma) = -\left\{\frac{\mu^2}{2\sigma^2} + \log \sigma + \log(\Phi_\beta - \Phi_\alpha)\right\}, \qquad \ldots \text{(2.13)}$$

and

$$-\begin{bmatrix} \dfrac{\partial u_1}{\partial \mu} & \dfrac{\partial u_2}{\partial \mu} \\[2mm] \dfrac{\partial u_1}{\partial \sigma} & \dfrac{\partial u_2}{\partial \sigma} \end{bmatrix}^{-1} \begin{bmatrix} \dfrac{\partial B}{\partial \mu} \\[2mm] \dfrac{\partial B}{\partial \sigma} \end{bmatrix} = \begin{bmatrix} \mu^2 + \sigma^2 \left\{1 - \dfrac{\left(t_\beta + \dfrac{2\mu}{\sigma}\right)Q_\beta - \left(t_\alpha + \dfrac{2\mu}{\sigma}\right)Q_\alpha}{\Phi_\beta - \Phi_\alpha}\right\} \\[4mm] \mu - \sigma\dfrac{Q_\beta - Q_\alpha}{\Phi_\beta - \Phi_\alpha} \end{bmatrix}$$

$$= \begin{bmatrix} (\mu + t'\sigma)^2 \\ \mu + t''\sigma \end{bmatrix} = \begin{bmatrix} x'^2 \\ x'' \end{bmatrix}, \quad x_\alpha < x', x'' < x_\beta.$$

Thus we see that the generalised condition of Huzurbazar is satisfied in this case. Hence the likelihood equations $\dfrac{\partial \log L_T(\mu, \sigma)}{\partial \mu} = 0$ and $\dfrac{\partial \log L_T(\mu, \sigma)}{\partial \sigma} = 0$ have a unique solution for all samples of any size. This solution maximises the likelihood function absolutely.

3. Doubly censored samples : case (i)

A doubly censored sample of size $N = r + n + s$ be drawn from a normal population with mean μ and standard deviation unity; there being r observations in the left truncated region and s observations in the right truncated region. The likelihood function of such a sample is

$$L_c(\mu) = C_1 \cdot \Phi_\alpha^r \cdot (1 - \Phi_\beta)^s \cdot \exp\left\{ -\frac{1}{2} \sum_{i=1}^{n} (x_i - \mu)^2 \right\} \qquad \dots (3.1)$$

where C_1 is a constant. Hence, the maximum likelihood estimating equation for μ is

$$\frac{\partial \log L_c(\mu)}{\partial \mu} = -r \cdot \frac{Q_\alpha}{\Phi_\alpha} + s \cdot \frac{Q_\beta}{1 - \Phi_\beta} + \sum_{i=1}^{n} (x_i - \mu) = 0. \qquad \dots (3.2)$$

To know whether (3.2) has a solution let us see which sign $\dfrac{\partial \log L_c(\mu)}{\partial \mu}$ has for extreme values of μ. It can be seen that

$$\lim_{\mu \to -\infty} \frac{\partial \log L_c(\mu)}{\partial \mu} = \infty. \qquad \dots (3.3)$$

On the other hand, $$\lim_{\mu \to \infty} \frac{\partial \log L_c(\mu)}{\partial \mu} = -\infty. \qquad \dots (3.4)$$

Because $\dfrac{\partial \log L_c(\mu)}{\partial \mu}$ is continuous for all μ, it follows from (3.3) and (3.4) that the likelihood equation $\dfrac{\partial \log L_c(\mu)}{\partial \mu} = 0$ has at least one solution in the range $(-\infty, \infty)$ of μ.

Now we shall show that the likelihood equation (3.2) has one and only one solution and the likelihood function $L_c(\mu)$ has absolute maximum at that point for all samples of any size. Consider the second order derivative of $\log L_c(\mu)$ wrt μ,

$$\frac{\partial^2 \log L_c(\mu)}{\partial \mu^2} = -\left[n + r(Q_\alpha + t_\alpha \Phi_\alpha) \frac{Q_\alpha}{\Phi_\alpha^2} + s\{Q_\beta - t_\beta(1 - \Phi_\beta)\} \frac{Q_\beta}{(1 - \Phi_\beta)^2} \right] \qquad \dots (3.5)$$

Since $$\frac{\partial(Q_\alpha + t_\alpha \Phi_\alpha)}{\partial t_\alpha} = \Phi_\alpha > 0,$$

and $$\lim_{t_\alpha \to -\infty} (Q_\alpha + t_\alpha \Phi_\alpha) = 0,$$

it follows that $$Q_\alpha + t_\alpha \Phi_\alpha > 0 \quad \text{for all } t_\alpha. \qquad \dots (3.6)$$

Again, since $$\frac{\partial\{Q_\beta - t_\beta(1 - \Phi_\beta)\}}{\partial t_\beta} = -(1 - \Phi_\beta) < 0,$$

and $$\lim_{t_\beta \to \infty} \{Q_\beta - t_\beta(1 - \Phi_\beta)\} = 0,$$

it also follows that $$Q_\beta - t_\beta(1 - \Phi_\beta) > 0, \quad \text{for all } t_\beta. \qquad \dots (3.7)$$

Hence $$\frac{\partial^2 \log L_c(\mu)}{\partial \mu^2} < 0 \quad \text{for all } \mu. \qquad \dots (3.8)$$

From this fact it follows that all the solutions of (3.2) provide maxima of the likelihood function. But, since $\log L_c(\mu)$ is continuous, between two maxima there should be a minimum for which

$$\frac{\partial^2 \log L_c(\mu)}{\partial \mu^2} > 0. \qquad \text{... (3.9)}$$

This contradicts (3.8). Hence, the likelihood equation (3.2) has a unique solution and that solution provides the absolute maximum of the likelihood function $L_c(\mu)$ for all samples of any size.

4. Doubly Censored Samples : Case (II)

Now, a censored sample of size $N = r+n+s$ be drawn from a normal population with mean zero and standard deviation σ. The likelihood function of the sample is

$$L_c(\sigma) = C_1 . \Phi_\alpha^r . (1-\Phi_\beta)^s \left(\frac{1}{\sigma}\right)^n \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} x_i^2 \right\} \qquad \text{... (4.1)}$$

and the likelihood equation for σ is

$$\frac{\partial \log L_c(\sigma)}{\partial \sigma} = \frac{1}{\sigma}\left[-n-r . \frac{t_\alpha Q_\alpha}{\Phi_\alpha} + s . \frac{t_\beta Q_\beta}{1-\Phi_\beta} + \frac{1}{\sigma^2} \sum_{i=1}^{n} x_i^2 \right] = 0. \qquad \text{... (4.2)}$$

Let us first see which sign $\dfrac{\partial \log L_c(\sigma)}{\partial \sigma}$ has for extreme values of σ, in order to know whether (4.2) has a solution. We can easily see that

$$\lim_{\sigma \to 0} \sigma \frac{\partial \log L_c(\sigma)}{\partial \sigma} = \infty; \qquad \text{... (4.3)}$$

and

$$\lim_{\sigma \to \infty} \sigma \frac{\partial \log L_c(\sigma)}{\partial \sigma} = -n < 0. \qquad \text{... (4.4)}$$

Since $\dfrac{\partial \log L_c(\sigma)}{\partial \sigma}$ is a continuous function of σ, from (4.3) and (4.4) we have that the likelihood equation (4.2) has at least one solution in the range $(0, \infty)$ of σ.

To establish uniqueness and maximum of the solution consider the second order derivative of $\log L_c(\sigma)$ wrt. σ. We have

$$\frac{\partial^2 \log L_c(\sigma)}{\partial \sigma^2} = -\frac{1}{\sigma^2} \frac{\partial \log L_c(\sigma)}{\partial \sigma} - \frac{1}{\sigma^2}\left[n+\sigma \frac{\partial \log L_c(\sigma)}{\partial \sigma} + r . \frac{t_\alpha^2 Q_\alpha}{\Phi_\alpha^2}(Q_\alpha+t_\alpha\Phi_\alpha) \right.$$
$$\left. + s\{Q_\beta-t_\beta(1-\Phi_\beta)\}\frac{t_\beta^2 Q_\beta}{(1-\Phi_\beta)^2} + \frac{1}{\sigma^2}\sum_{i=1}^{n} x_i^2 \right]. \qquad \text{... (4.5)}$$

For all the solutions of (4.2),

$$\frac{\partial^2 \log L_c(\sigma)}{\partial \sigma^2} = -\frac{1}{\sigma^2}\left[n+r(Q_\alpha+t_\alpha\Phi_\alpha)\frac{t_\alpha^2 Q_\alpha}{\Phi_\alpha^2} + s\{Q_\beta-t_\beta(1-\Phi_\beta)\}\frac{t_\beta^2 Q_\beta}{(1-\Phi_\beta)^2} + \frac{1}{\sigma^2}\sum_{i=1}^{n} x_i^2 \right] < 0.$$
$$\text{... (4.6)}$$

As in Section 3 we can easily see that this leads to contradiction if (4.2) has more than one solution. Hence, the likelihood equation $\dfrac{\partial \log L_c(\sigma)}{\partial \sigma} = 0$ has one and only one solution for all samples of any size, and this solution does make the likelihood function $L_c(\sigma)$ maximum.

5. DOUBLY CENSORED SAMPLES : CASE (III)

In this case, a doubly censored sample of size $N = r+n+s$ is drawn from a normal population with mean μ and standard deviation σ. The likelihood function of the sample is

$$L_c(\mu, \sigma) = C_1 \cdot \Phi_\alpha^r (1-\Phi_\beta)^s \left(\frac{1}{\sigma}\right)^n \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i-\mu)^2\right\} \qquad \ldots \quad (5.1)$$

Now, for all the solutions of the likelihood equations $\dfrac{\partial \log L_c(\mu, \sigma)}{\partial \mu} = 0$

and

$$\frac{\partial \log L_c(\mu, \sigma)}{\partial \sigma} = 0$$

we have

$$-\sigma^2 \frac{\partial^2 \log L_c(\mu, \sigma)}{\partial \mu^2} = a_\alpha + a_\beta + n,$$

$$-\sigma^2 \frac{\partial^2 \log L_c(\mu, \sigma)}{\partial \mu \partial \sigma} = t_\alpha a_\alpha + t_\beta a_\beta + \frac{1}{\sigma} \sum_{i=1}^{n} (x_i-\mu), \qquad \ldots \quad (5.2)$$

and

$$-\sigma^2 \frac{\partial^2 \log L_c(\mu, \sigma)}{\partial \sigma^2} = t_\alpha^2 a_\alpha + t_\beta^2 a_\beta + n + \frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i-\mu)^2,$$

where

$$a_\alpha = r(Q_\alpha + t_\alpha \Phi_\alpha) \frac{Q_\alpha}{\Phi_\alpha^2}, \qquad \ldots \quad (5.3)$$

$$a_\beta = s\{Q_\beta - t_\beta(1-\Phi_\beta)\} \frac{Q_\beta}{(1-\Phi_\beta)^2}.$$

Hence

$$\sigma^2 \begin{bmatrix} -\dfrac{\partial^2 \log L_c(\mu, \sigma)}{\partial \mu^2} & -\dfrac{\partial^2 \log L_c(\mu, \sigma)}{\partial \mu \partial \sigma} \\ \\ -\dfrac{\partial^2 \log L_c(\mu, \sigma)}{\partial \mu \partial \sigma} & -\dfrac{\partial^2 \log L_c(\mu, \sigma)}{\partial \sigma^2} \end{bmatrix}$$

$$= \begin{bmatrix} a_\alpha + a_\beta & t_\alpha a_\alpha + t_\beta a_\beta \\ \\ t_\alpha a_\alpha + t_\beta a_\beta & t_\alpha^2 a_\alpha + t_\beta^2 a_\beta \end{bmatrix}$$

$$+ \sum_{i=1}^{n} \begin{bmatrix} 1 & \dfrac{x_i-\mu}{\sigma} \\ \\ \dfrac{x_i-\mu}{\sigma} & \dfrac{(x_i-\mu)^2}{\sigma^2} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & n \end{bmatrix},$$

$$= \begin{bmatrix} \sqrt{a_\alpha} & \sqrt{a_\beta} \\ t_\alpha \sqrt{a_\alpha} & t_\beta \sqrt{a_\beta} \end{bmatrix} \cdot \begin{bmatrix} \sqrt{a_\alpha} & t_\alpha \sqrt{a_\alpha} \\ \sqrt{a_\beta} & t_\beta \sqrt{a_\beta} \end{bmatrix}$$

$$+ \sum_{i=1}^{n} \begin{bmatrix} 1 \\ \dfrac{x_i-\mu}{\sigma} \end{bmatrix} \cdot \begin{bmatrix} 1 & \dfrac{x_i-\mu}{\sigma} \end{bmatrix} + \begin{bmatrix} 0 \\ \sqrt{n} \end{bmatrix} \cdot \begin{bmatrix} 0 & \sqrt{n} \end{bmatrix} \qquad \ldots \quad (5.4)$$

The right hand side of (5.4) is a sum of Gram matrices; and the rank of first matrix is 2 and that of other matrices is unity. Hence, it follows that the matrix on the left side of (5.4) is positive definite. Therefore it is evident that a solution of the likelihood equations $\dfrac{\partial \log L_c(\mu, \sigma)}{\partial \mu} = 0$ and $\dfrac{\partial \log L_c(\mu, \sigma)}{\partial \sigma} = 0$ is unique, and the likelihood function $L_c(\mu, \sigma)$ is maximum at this point for every sample of any size.

References

COHEN, A. C., JR. (1950): Estimating the mean and variance of normal populations from singly and doubly truncated samples. *Ann. Math. Stat.*, **21**, 557-569.

———— (1959): Simplified estimators for the normal distribution when samples are singly censored or truncated. *Technometrics*, **1**, 217-237.

DES RAJ (1954): Truncated sampling from distributions admitting sufficient statistics. *Sankhyā*, **14**, 169-174.

GUPTA, A. K. (1952): Estimation of the mean and standard deviation of a normal population from a censored sample. *Biometrika*, **39**, 260-273.

HALD, A. (1949): Maximum likelihood estimation of the parameters of a normal distribution which is truncated at a known point. *Skand. Aktuar Tidskr.*, **32**, 119-134.

HUZURBAZAR, V. S. (1948): The likelihood equation, consistency and the maxima of the likelihood function. *Ann. Eugen.*, **14**, 185-200.

———— (1949): On a property of distributions admitting sufficient statistics. *Biometrika*, **36**, 71-74.

KOOPMAN, B. O. (1936): On distributions admitting sufficient statistic. *Trans. Amer. Math. Soc.*, **39**, 399-409.

PITMAN, E. T. G. (1936): Sufficient statistics and intrinsic accuracy. *Proc. Camb. Phil. Soc.*, **32**, 567-579.

TUKEY, J. W. (1949): Sufficiency, truncation and selection. *Ann. Math. Stat.*, **20**, 309-311.

THE PROBLEM OF TESTING LINEAR HYPOTHESIS ABOUT POPULATION MEANS WHEN THE POPULATION VARIANCES ARE NOT EQUAL AND M-TEST

By SAIBAL BANERJEE

Indian Statistical Institute

SUMMARY. Given k independent samples of n_i units from k populations $N(m_i, \sigma_i^2)$ ($i = 1$, $2, ..., k$) a test statistic for testing a hypothesis H_0 about s ($s \leqslant k$) linear functions of k population means without any *a priori* knowledge of population variances or the ratio of the variances is of interest. A new test statistic called M statistic is defined for testing such hypothesis where any prior knowledge about the population variances is not available. The error of the first kind (probability of rejection of the hypothesis when true) of the test statistic depends on the unknown population variances but the test statistic is so defined that for all possible values of population variances the error of the first kind is less than or equal to some pro-assigned probability α. It is shown that critical values of the test statistic for testing a hypothesis about two linear functions of k population means with $\alpha = 0.05, 0.02, 0.01$, etc., can all be obtained from tabulated values of F-table. A numerical example for testing equality of three population means has been considered. It is also shown that the test statistic can be used in multivariate problems as well. An analysis of Barnard's data (Barnard, 1935) has been considered.

1. INTRODUCTION

1.1. Given k samples of n_i units from k normal populations $N_i(m_i, \sigma_i^2)$ ($i = 1, 2, ..., k$) having equal variances or the ratio of the variances known *a priori* any hypothesis about any linear function $\sum_1^k c_i m_i$ of population means (where c_i ($i = 1, 2, ..., k$) are known coefficients) can be tested by the t-statistic. Also, any hypothesis about more than one linear function of population means can be tested by F-statistic or F-ratio. If the population variances are not equal or the ratio of the variances are not known *a priori* it is possible to test (Banerjee, 1961) any hypothesis about any single linear function of population means. Also, any hypothesis about more than one linear function of population means can be tested by a new statistic hereinafter called M-statistic or M-ratio.

2. SAMPLES FROM HETEROSCEDASTIC POPULATIONS

2.1. Let x_i, s_i^2 ($i = 1, 2, ..., k$) be sample estimates of population means and variances of k samples of n_i-units drawn drom k normal population $N_i(m_i, \sigma_i^2)$ ($i = 1, 2, ..., k$). Suppose it is required to test the hypothesis that

$$H_0 \begin{cases} c_{11}m_1 + c_{12}m_2 + ... + c_{1k}m_k = M_1 & ... \quad (2.1.1) \\ c_{21}m_1 + c_{22}m_2 + ... + c_{2k}m_k = M_2 & ... \quad (2.1.2) \\ \vdots \\ c_{s1}m_1 + c_{s2}m_2 + ... + c_{sk}m_k = M_s & ... \quad (2.1.s) \end{cases}$$

where c_{ij} $(i = 1, 2, ..., s; j = 1, 2, ..., k)$ and M_j $(j = 1, 2, ..., s)$ are known constants. It is assumed without any loss of generality that the relations (2.1.1), (2.1.2), ... (2.1.3) are mutually consistent and independent. It is also assumed that $s < k$ for if $s = k$ the relations (2.1.1), (2.1.2), ... (2.1.3) can be replaced by

$$m_i = M_i' \quad (i = 1, 2, ..., k)$$

and H_0 can be tested by the statistic

$$T = \sum_1^k \left[\frac{x_i - M_i'}{s_i \sqrt{n}} \right]^2 = \sum_1^k t_i^2$$

whose percentage points, although not tabulated, can be evaluated as each t_i $(i = 1, 2, ..., k)$ would be independently distributed as a Student's t-variate if the hypothesis be true.

2.2. Let test variates $U_1, U_2, ..., U_s$ be defined as

$$U_i = \sum_{j=1} c_{ij} \bar{x}_j. \quad (i = 1, 2, ..., s). \qquad ... \quad (2.2.1)$$

The test variates $U_1, U_2, ..., U_s$ as defined in (2.2.1) are stochastic variates jointly distributed in a multivariate normal form.

2.3. Now let us consider the probability of the inequality

$$\sum_{i=1}^s (U_i - M_i)^2 > \sum_{j=1}^k A_j C_j \frac{\sigma_i^2}{n_j}$$

where $C_1, C_2, ..., C_k$ are defined as

$$C_j = \sum_{i=1}^s c_{ij}^2; \quad (j = 1, 2, ..., k)$$

and A_j $(j = 1, 2, ..., k)$ are positive constants to be suitably determined in a manner as discussed later.

2.4. Let $M_1', M_2', ..., M_s'$ be respectively means of test variates $U_1, U_2, ..., U_s$ whereas by hypothesis H_0 the means are $M_1, M_2, ..., M_s$. Let variates u_i $(i = 1, 2, ..., s)$ be defined to

$$u_i = U_i - M_i'; \quad (i = 1, 2, ..., s) \qquad ... \quad (2.4.1)$$

u_i $(i = 1, 2, ..., s)$ as defined in (2.4.1) follow a multivariate normal distribution with zero mean with, say, disperson matrix R. Now consider a further transformation (Ferrar, 1953) to variates v_i $(i = 1, 2, ..., s)$ so that

$$\left. \begin{array}{c} \sum_1^s u_i^2 = \sum_1^s v_i^2 \\[2mm] uR^{-1}u' = \lambda_1 v_1^2 + \lambda_2 v_2^2 + ... + \lambda_s v_s^2 \end{array} \right\} \qquad ... \quad (2.4.2)$$

and

where

u is a row matrix $(u_1, u_2, ..., u_s)$,

u' is a transpose u

and

R^{-1} is a $s \times s$ matrix reciprocal to the dispersion matrix R.

The transformed variates v_i $(i = 1, 2, ..., s)$ are independently normally distributed with zero mean and variance, say, σ_{vi}^2 $(i = 1, 2, ..., s)$.

2.5. Now by virtue of (2.4.1) and (2.4.2)

$$\sum_1^s (U_i - M_i)^2 = \sum_1^s (u_i - M_i + M_i')^2 = \sum_1^s (u_i - d_i)^2 = \sum_1^s (v_i - d_i')^2 \quad \dots \quad (2.5.1)$$

$$\left(\text{where } \sum_1^s d_i^2 = \sum_1^s d_i'^2 \right).$$

Also, by virtue of (2.4.1)

$$\sum_1^s V(U_i) = \sum_1^s V(u_i) = \sum_1^s E(u_i^2) = \sum_1^s E(v_i^2) = \sum_1^s \sigma_{v_i}^2. \quad \dots \quad (2.5.2)$$

2.6. From (2.5.1) and (2.5.2) the probability of the inequality

$$\sum_1^s (U_i - M_i)^2 > \sum_1^k A_j C_j \frac{s_j^2}{n_j}$$

is equal to

$$\frac{\sum_1^s (v_i - d_i)^2}{\sum_1^s \sigma_{v_i}^2} \left\{ = \frac{\sum_{i=1}^s (U_i - M_i)^2}{\sum_{i=1}^s V(U_i)} \right\} \geqslant \frac{\sum_{j=1}^k A_j C_j \frac{s_j^2}{n_j}}{\sum_i V(U_i)}$$

which is equal to

$$\sum_{i=1}^s \beta_i \chi_{1i}^2 \geqslant \sum_{j=1}^k A_j \omega_j \frac{\chi_j^2}{\nu_j}$$

where

χ_{1i}^2 are non-central χ^2-variates with 1 d.f. $(i = 1, 2, \dots, s)$

χ_j^2 are χ^2—variates with ν_j d.f. $(\nu_j = n_j - 1)$, $(j = 1, 2, \dots, k)$

β_j and ω_j are positive weights defined as

$$\beta_i = \frac{\sigma_{v_i}^2}{\sum_1^s \sigma_{v_i}^2} ; \omega_j = \frac{C_j^2 \sigma_j^2 / n_j}{\sum_1^k C_j^2 \sigma_j^2 / n_j}. \quad (i = 1, 2, \dots, s; \ j = 1, 2, \dots, k)$$

If the hypothesis H_0 is true $\chi_{1i}^2 (i = 1, 2, \dots, s)$ are, however, distributed as χ^2-variates with 1 d.f.

2.7. The crux of the problem of having a test statistic for testing hypothesis H_0 based on test variates U_i $(i = 1, 2, \dots, s)$ therefore boils down to finding positive constants A_j $(j = 1, 2, \dots, k)$ so that

$$\text{prob} \left[\sum_{i=1}^s \beta_i \chi_{1i}^2 > \sum_{j=1}^k A_j \omega_j \frac{\chi_j^2}{\nu_j} \right] < \alpha \quad \dots \quad (2.7.1)$$

where χ_{1i}^2 $(i = 1, 2, \dots, s)$ and χ_j^2 $(j = 1, 2, \dots, k)$ are all independently distributed χ^2-variates with respectively 1 and ν_j $(j = 1, 2, \dots, k)$ d.f. and β_i and ω_j are positive weights adding up to unity. First, it has, however, to be proved that it is *at all possible* to find finite positive constants A_j $(j = 1, 2, \dots, k)$ so that given some pre-assigned α (2.8.1) would be satisfied.

2.8. Theorem : *Let U_i $(i = 1, 2, \dots, s)$ be a set of stochastic variates (not necessarily independently distributed) which satisfy the relation that*

$$prob \, [U_i \leqslant 0] \leqslant \alpha_i \qquad (i = 1, 2, \dots, s). \quad \dots \quad (2.8.1)$$

Now if β_i $(i = 1, 2, ..., s)$ be a set of arbitrary positive weights adding up to unity (i.e. $\sum_1^s \beta_i = 1$), then

$$prob \; [\sum_1^s \beta_i U_i < 0] < \sum_1^s \alpha . \qquad \qquad ... \; (2.8.1)$$

Proof: First, let us consider the case of only two variates — U_1 and U_2. Now if β_1 and β_2 be two positive weights adding up to unity

$$prob \; [\beta_1 u_1 + \beta_2 u_2 \leqslant 0] \leqslant prob \; [U_1 \leqslant 0] + prob \; [U_2 \leqslant 0] \leqslant \alpha_1 + \alpha_2.$$

Also, similarly it can be proved that

$$prob \; [\sum_{i=1}^s \beta_i U_i < 0] < \sum_{i=1}^s \alpha_i, \qquad \qquad ... \; (2.8.2)$$

2.9. Now let U_i $(i = 1, 2, ..., s)$ be defined as

$$\sum_{j=1}^s A_j \omega_j \frac{\chi_j^2}{\nu_j} - \chi_{1i}^2 \qquad \qquad (i = 1, 2, ..., s) \qquad \qquad ... \; (2.9.1)$$

where χ_{1i}^2 $(i, 1. 2. ..., s)$ and χ_j^2 $(j = 1, 2, ..., k)$ are all independently distributed χ^2-variates with respectively 1 and ν_j $(j = 1, 2, ..., k)$ d.f. and A_j $(j = 1, 2, ..., k)$ are 100. α/s percentile point of Student's t-table of d.f. ν_j $(j = 1, 2, ..., k)$ so that (Banerjee, 1960)

$$prob \; [U_i \leqslant 0] \leqslant \alpha/s. \qquad \qquad ... \; (2.9.2)$$

From (2.8.1) and (2.8.2) it follows

$$prob \; [\sum_1^s \beta_i \chi_{1i}^2 \geqslant \sum_1^s A_j \omega_j \chi_j^2/\nu_j] < \alpha. \qquad \qquad ... \; (2.9.3)$$

3. STATEMENT OF THE STATISTIC

3.1. Let $M_{s,t}$-statistic (M after Mahalanobis) for testing hypothesis about s linear functions of population means without any *a priori* knowledge of population variances of size α (or with maximum value of error of the first kind α) be defined as

$$\cfrac{\sum_1^s \beta_i \chi_{1i}^2}{\sum_{j=1}^s A_j \omega_j \frac{\chi_j^2}{\nu_j}}$$

where χ_{1i}^2 $(i = 1, 2, ..., s)$ and χ_j^2 $(j = 1, 2, ..., k)$ are independently distributed χ^2-variates with respectively 1 and ν_j $(j = 1, 2, ..., k)$ d.f. and β_i and ω_j $(i = 1, 2, ..., s; j = 1, 2, ..., k)$ are a set of positive weights adding up to unity and A_s $(j = 1, 2, ..., k)$ are irreducible positive constants which have been so determined so that

$$prob \left[\; [\sum_1^s \beta_i \chi_{1i}^2 > \sum_1^s A_j \omega_j \frac{\chi_j^2}{\nu_j} \right]$$

is less than or equal to α for all possible values of β_i and ω_j $(i = 1, 2, ..., s; j = 1, 2, ..., k)$.

4. CRITICAL VALUES OF M-STATISTIC

4.1. Let us consider the case of finding critical values of M-statistic for the case $s = 2$ and any k. The problem of finding critical values of $M_{2,k}$ amounts to finding minimum possible numerical values of A_j $(j = 1, 2, ..., k)$ so that

$$\text{prob}\left[\sum_1^2 \beta_i \chi_{1i}^2 \geqslant \sum_1^k A_j \omega_j \frac{\chi_j^2}{\nu_j}\right] \leqslant \alpha. \qquad \text{... (4.1.1)}$$

If P denotes the probability of the inequality

$$\sum_1^2 \beta_i \chi_{1i}^2 > \sum_1^k A_j \omega_j \frac{\chi_j^2}{\nu_j} \qquad \text{... (4.1.2)}$$

we have

$$1 - P = \int_0^\infty \int_0^\infty \cdots \int_0^\infty \prod_{i=1}^k f(\chi_j^2)\left[\int_0^{T/\beta_1} h(\chi_{11}^2)\left\{\int_0^{(T-\beta_1\chi_{11}^2)/\beta_2} h(\chi_{12}^2)d\chi_{12}^2\right\}d\chi_{11}^2\right]d\chi_1^2\chi_2^2...d\chi_k^2... \quad \text{(4.1.3)}$$

where $h(\chi_{1i}^2)$ $(i = 1, 2)$ denotes frequency function of a χ^2-variate with 1d.f. $(i = 1, 2)$ $f(\chi_j^2)$ denotes frequency functions of χ^2-variate with ν_j d.f. $(j = 1, 2, ..., k)$ and
$$T = \sum_{j=1}^k A_j \omega_j \frac{\chi_j^2}{\nu_j}.$$

4.2. The integral $\int_0^{z/\beta_1} h(\chi_{11}^2)\left\{\int_0^{z(-\beta_1\chi_{11}^2)/\beta_2} h(\chi_{12}^2)d\chi_{12}^2\right\} d\chi_{11}^2$ is an upward convex function of z (Courant, 1957) (details in Appendix A.1) so that

$$\int_0^{\sum_1^k \omega_i z_i/\beta_1} h(\chi_{11}^2)\left\{\int_0^{(\sum_1^k \omega_i z_i - \beta_1\chi_{11}^2)/\beta_2} h(\chi_{12}^2)d\chi_{12}^2\right\}d\chi_{11}^2$$

$$\geqslant \sum_1^k \omega_i \int_0^{z_i/\beta_1} h(\chi_{11}^2)\left\{\int_0^{(z_i-\beta_1\chi_{11}^2)/\beta_1} h(\chi_{12}^2)d\chi_{12}^2\right\}d\chi_{11}^2. \qquad \text{... (4.2.1)}$$

From (4.1.1), (4.1.2) and (4.1.3) it follows

$$P < \sum_{j=1}^k \omega_j P_j \qquad \text{... (4.2.2)}$$

where
$$P_j = \int_0^\infty f(\chi_j^2)\left[\int_{T_1/\beta_1}^\infty h(\chi_{11}^2)\left\{\int_{(T_1-\beta_1\chi_{11}^2)/\beta_2}^\infty h(\chi_{12}^2)d\chi_1^2.\right\}d\chi_{11}^2\right]d\chi_j^2$$

$$= \int_0^\infty \int_0^\infty h(\chi_{11}^2)h(\chi_{12}^2)\left\{\int_0^{T_2} f(\chi_j^2)d\chi_j^2\right\}d\chi_{11}^2 d\chi_{12}^2 \qquad \text{... (4.2.3)}$$

where
$$T_1 = \frac{A_j\chi_j^2}{\nu_j}$$

and
$$T_2 = \frac{\beta_1\chi_{11}^2 + \beta_2\chi_{12}^2}{\dfrac{A_j}{\nu_j}}.$$

4.3. Now, for degrees of freedom of χ_j^2 equal to 1 or 2, the integral

$$\int_0^\infty \int_0^\infty h(\chi_{11}^2)h(\chi_{12}^2)\left\{\int_0^{T_3} f(\chi^2)d\chi_j^2\right\}d\chi_{11}^2 d\chi_{12}^2 \qquad \ldots \quad (4.3.1)$$

where

$$T_3 = \frac{\beta_1\chi_{11}^2 + \beta_2\chi_{12}^2}{F_{2,\nu_j.\alpha}/\nu_j}$$

for variation in β_1 and β_2 is always less than or equal to α, where $F_{2,\nu_j.\alpha}$ is tabulated F-value of F-table corresponding to 100α percentage point and d.f. of greater mean square 2 and d.f. of smaller mean square ($\nu_j = 1,2$). (Details in Appendix A.2).

4.4. Also, for the case $\nu_j \geqslant 3$ and $\alpha = 0.05, 0.02, 0.01$, etc., the integral

$$\int_0^\infty \int_0^\infty h(\chi_{11}^2)\, h(\chi_{12}^2)\{\int_0^{T_4} f(\chi_j^2)d\chi_j^2\}\, d\chi_{11}^2.d\chi_{12}^2 \qquad \ldots \quad (4.4.1)$$

where

$$T_4 = \frac{\beta_1\chi_{11}^2 + \beta_2\chi_{12}^2}{F_{1,\nu_j.\alpha}/\nu_j}$$

for variation in β_1 and β_2 is always less than or equal to α, where $F_{1,\nu_j.\alpha}$ is tabulated F-value of F-table corresponding to 100α percentage point and d.f. of greater mean square 1 and d.f. of smaller mean square ($\nu_j \geqslant 3$). (Details in Appendix A.2).

4.5. Numerical values of A_j of $M_{2,k}$ test can thus be determined from tabulated values F-table. Table 1 below gives numerical values of A_j of $M_{2,k}$ test of size 0.05 and d.f. $\nu_j = 1, 2, \ldots, 20$. The values have been taken from F-table.

TABLE 1. NUMERICAL VALUES OF
A_j OF M_{2k} TEST OF SIZE 0.05

ν_j	A_j	ν_j	A_j
1	200.00	11	4.84
2	19.00	12	4.75
3	10.13	13	4.67
4	7.71	14	4.60
5	6.61	15	4.54
6	5.99	16	4.49
7	5.59	16	4.45
8	5.32	18	4.41
9	5.12	19	4.38
10	4.96	20	4.35

5. TESTING EQUALITY OF POPULATION MEANS

5.1.   Given k samples from k normal populations $N_i(m_i, \sigma_i^2)$ to test the equality of population means $k-1$ mutually independent linear functions $L_i\ (i = 1, 2, ..., k-1)$ of population means and associated test variates may be defined as

$$L_i = \sum_{j=1}^{k} c_{ij} m_j\ ;\qquad U_i = \sum_{j=1}^{k} c_{ij} \cdot x_j\ ;\qquad \dots \text{(5.1.1)}$$

$$(i = 1, 2, ..., k-1).$$

where $\sum_{j=1}^{k} c_{ij} = 0$.   If s_i^2 denotes estimate of population variance of the i-th population $(i = 1, 2, ..., k)$ $M_{k-1,k}$-statistic may be computed as

$$\dfrac{\sum\limits_{i=1}^{k-1} U_i^2}{\sum\limits_{j=1}^{k} A_j C_j \dfrac{s_j^2}{n_j}} \qquad \dots \text{(5.1.2)}$$

(where $C_j = \sum_{i=1}^{k} c_{ij}^2$; $j = 1, 2, ..., k$) with suitable choice of $A_j\ (j = 1, 2, ..., k)$ and the hypothesis would be rejected if the numerical value of $M_{k-1,k}$ as defined in (5.1.2) exceeded unity.

6. NUMERICAL EXAMPLE

6.1.   Three samples from three populations supply the following estimates.

TABLE 2

		population		
		I	II	II
sample mean	x_i	5.0	20.0	10.0
sample variance	s_i^2	18.0	5.5	20.0
sample size	n_i	3	11	21

Defining test variates U_1, and U_2 as

$$U_1 = \frac{1}{\sqrt{2}}(\bar{x}_1 - x_2) = \frac{1}{\sqrt{2}}(5-20)$$

$$U_2 = \frac{1}{\sqrt{6}}(\bar{x}_1 + x_2 - 2x_3) = \frac{1}{\sqrt{6}}(25-20).$$

$M_{2\cdot3}$-statistic of size .05 may be computed as

$$M_{2\cdot3} = \frac{U_1^2 + U_2^2}{\frac{2}{3}\left[\frac{A_1 s_1^2}{n_1} + \frac{A_2 s_2^2}{n_2} + \frac{A_3 s_3^2}{n_3}\right]}$$

$$= \frac{\frac{1}{2}[5-20]^2 + \frac{1}{6}[25-20]^2}{\frac{2}{3}\left[19.00 \times \frac{18}{3} + 4.96 \times \frac{5.5}{11} + 4.35 \times \frac{20}{21}\right]}$$

$$= \frac{\frac{225}{2} + \frac{25}{6}}{\frac{2}{3}[114.00 + 2.48 + 4.14]}$$

$$= \frac{116.67}{80.41} = 1.45$$

where numerical values of $A_j (j = 1, 2, 3)$ have been taken from Table 1 above. Since $M_{2\cdot3}$ is greater than unity any hypothesis about equality of means is rejected.

7. THE CASE OF MULTIVARIATE POPULATION

7.1. Let k samples of N_i $(i = 1, 2, ..., k)$ units be drawn from k, p-variate normal populations having dispersion matrices Σ_i $(i = 1, 2, ..., k)$ which are not necessarily equal. Let $\bar{x}_{ij}$ and m_{ij} denote sample mean and population mean of j-th character of i-th population. Also let s_{ij} and σ_{ij} denote sample and population variance of j-th character of i-th population. To test the hypothesis that

$$\sum_{i=1}^{k} c_{ij} m_{ij} = \lambda_j \qquad (j = 1, 2, ..., p). \qquad \qquad \ldots \ (7.1.1)$$

$M_{p,pk}$-statistic may be defined as

$$\frac{\sum_{j=1}^{p}\left\{\sum_{i=1}^{k} c_{ij}\bar{x}_{ij} - \lambda_j\right\}^2}{\sum_{i=1}^{k}\frac{A_i}{N_i}\sum_{j=1}^{p} c_{ij}^2 s_{ij}^2} \qquad \qquad \ldots \ (7.1.2)$$

with suitable choice of A_i $(i = 1, 2, ..., k)$ depending upon the size of the test. It can be shown that $M_{p,pk}$ as defined in (7.1.2) is equal to

$$\frac{\sum_{1}^{p} \beta_i \chi_{1i}^2}{\sum_{i=1}^{k} A_i \sum_{j=1}^{p} \omega_{ij}\frac{\chi_{ij}^2}{\nu_{ij}}} \qquad \qquad \ldots \ (7.1.3)$$

where χ_{1i}^2 $(i = 1, 2, ..., p)$ and χ_{ij}^2 $(i = 1, 2, ..., k; j = 1, 2, ..., p)$ are independently distributed χ^2-variates, χ_{1i}^2 being distributed with 1 and χ_{ij}^2 being distributed with $N_i - 1$ d.f. and β_i and ω_{ij} are a set of positive weights adding up to unity i.e. $\sum_{i=1}^{k} \beta_i = 1$ and $\sum_{i=1}^{k}\sum_{j=1}^{p} \omega_{ij} = 1$.

8. FURTHER NUMERICAL EXAMPLE

8.1. As an example of likely use of M-statistic in multivariate problems let us consider Barnard's data on Egyptian skulls. Four measurements on four populations are summarised as

TABLE 3. MEAN VALUES OF FOUR CHARACTERS

	character			
	I	IV	VI	VII
population I	133.582	98.308	50.835	133.000
II	134.265	96.463	51.148	134.883
III	134.371	95.857	50.100	133.643
IV	135.307	95.040	52.093	131.467

with numbers of observations as $N_1 = 91$, $N_2 = 162$, $N_3 = 70$ and $N_4 = 75$ and pooled corrected sum of squares of the four characters as (i) 9661.997, (ii) 9073.115, (iii) 3938.320 and (iv) 8741.509. Let $\bar{x}_{ij}$ and m_{ij} denote sample mean and population mean of j-th character of the i-th population $(i, j = 1, 2, 3, 4)$. Also let s_j^2 and σ_j^2 denote sample and population variances of the j-th character. (Here the dispersion matrices of the populations have been assumed to be equal.) To test the hypothesis that

$$m_{1j} = m_{2j} = m_{3j} = m_{4j} \quad (j = 1, 2, 3, 4).$$

Let test variates U_{jk} $(j = 1, 2, 3, 4; \quad k = 1, 2, 3)$ be defined be

$$U_{j1} = \frac{1}{\sqrt{2}} \{\bar{x}_{1j} - \bar{x}_{2j}\}$$

$$U_{j2} = \frac{1}{\sqrt{2}} \{\bar{x}_{3j} - \bar{x}_{4j}\} \qquad \left.\right\} \quad \ldots \quad (8.1.1)$$

$$U_{j3} = \frac{1}{\sqrt{4}} \{\bar{x}_{1j} + \bar{x}_{2j} - \bar{x}_{3j} - \bar{x}_{4j}\}$$

$$(j = 1, 2, 3, 4).$$

On the basis of test variates U_{jk} $(j = 1, 2, 3, 4; \quad k = 1, 2, 3)$ $M_{12,4}$-statistic may be computed as

$$\frac{\sum\limits_{j=1}^{4} \sum\limits_{k=1}^{3} U_{jk}^2}{\frac{3}{4} A \sum\limits_{j=1}^{4} s_j^2 \left\{ \frac{1}{N_1} + \frac{1}{N_2} + \frac{1}{N_3} + \frac{1}{N_4} \right\}} \qquad \ldots \quad (8.1.2)$$

with suitable choice of A depending on the size of the test. Taking numerical value of A equal to 3.86 (value taken from tabulated 5 p.c. point of F-table corresponding to $\nu_1 = 1$ and $\nu_2 = 400$) *approximate* numerical value of $M_{12,4}$-statistic comes out as 1.49 . Since numerical value of $M_{12,4}$-statistic exceeds unity the hypothesis cannot be accepted.

6

Appendix A.1

Let

$$F(z) \equiv \int_0^{z/\beta_1} e^{-x_1} x_1^{-\frac{1}{2}} \left\{ \int_0^{(z-\beta_1 x_1)\beta_2} e^{-x_2} x_2^{-\frac{1}{2}} dx_2 \right\} ds_1 \qquad \dots \text{(A.1.1)}$$

where $\beta_1 + \beta_2 = 1 \; ; \; \beta_1, \beta_2 > 0$ and $\beta_2 > \beta_1$.

We have

$$\frac{d}{dz} F(z) = \int_0^{z/\beta_1} e^{-x_1} x_1^{-\frac{1}{2}} e^{-(z-\beta_1 x_1)/(1-\beta_1)} \left\{ \frac{z-\beta_1 x_1}{1-\beta_1} \right\}^{-\frac{1}{2}} \left\{ \frac{1}{1-\beta_1} \right\} dx_1$$

$$= K. \int_0^z e^{-x/\beta_1} e^{-(z-x)/(1-\beta_1)} \left\{ z-x \right\}^{-\frac{1}{2}} x^{-\frac{1}{2}} dx = I_1 + I_2 \qquad \dots \text{(A.1.2)}$$

where

$$I_1 = K. e^{-x/\beta_1} e^{-(z-x)/(1-\beta)} \int x^{-\frac{1}{2}} (z-x)^{-\frac{1}{2}} dx \Big]_0^z$$

$$= 2K. e^{-x/\beta_1} e^{-(z-x)/(1-\beta_1)} \sin^{-1} \sqrt{\frac{x}{z}} \Big]_0^z$$

$$= 2K. e^{-z/\beta_1} \frac{\pi}{2} = K. e^{-z/\beta_1} \qquad \dots \text{(A.1.3)}$$

and

$$I_2 = -K \int_0^z e^{-x/\beta_1} e^{-(z-x)/(1-\beta_1)} \left\{ -\left(\frac{1}{\beta_1} - \frac{1}{1-\beta_1} \right) \right\} \times 2 \sin \sqrt{\frac{x}{z}} dx. \qquad \dots \text{(A.1.4)}$$

Now,

$$\frac{d}{dz} I_1 = \pi e^{-z/\beta_1} \left\{ -\frac{1}{\beta_1} \right\} \qquad \dots \text{(A.1.5)}$$

and

$$\frac{d}{dz} I_2 = I_{21} + I_{22} \qquad \dots \text{(A.1.6)}$$

where

$$I_{21} = 2K e^{-z/\beta_1} \left\{ \frac{1}{\beta_1} - \frac{1}{1-\beta_1} \right\} \frac{\pi}{2}$$

$$= K. e^{-z/\beta} \left\{ \frac{1}{\beta_1} - \frac{1}{1-\beta_1} \right\} \pi$$

and

$$I_{22} = K \int_0^z e^{-x/\beta_1} \left\{ \frac{1}{\beta_1} - \frac{1}{1-\beta_1} \right\} e^{-x/(1-\beta_1)}$$

$$\times \frac{d}{dz} \left[e^{-z/(1-\beta_1)} \sin^{-1} \sqrt{\frac{x}{z}} \right] dx. \qquad \dots \text{(A.1.7)}$$

As

$$\frac{d}{dz} \left\{ e^{-z/(1-\beta_1)} \sin^{-1} \sqrt{\frac{x}{z}} \right\} = e^{-z/(1-\beta_1)} \sin^{-1} \sqrt{\frac{x}{z}} \left\{ -\frac{1}{1-\beta_1} \right\}$$

$$+ e^{-z/(1-\beta_1)} \left\{ \frac{1}{2} \frac{z}{\sqrt{x} \sqrt{z-x}} \left(-\frac{x}{z^2} \right) \right\},$$

$$I_{22} = K \int_0^z e^{-x/\beta_1} e^{-(z-x)/(1-\beta_1)} \left\{ \frac{1}{\beta_1} - \frac{1}{1-\beta_1} \right\} \times \left[\sin^{-1} \sqrt{\frac{x}{z}} \left(-\frac{1}{1-\beta_1} \right) - \frac{1}{2z} \frac{x^{\frac{1}{2}}}{\sqrt{z-x}} \right] dx$$

$$\dots \text{(A.1.8)}$$

From (A.1.1), (A.1.2), ... (A.1.8) it follows

$$\frac{d^2}{dz^2} F(z) = -K e^{-z/\beta_1} \left\{ \frac{1}{1-\beta_1} \right\} + I_{22}.$$

As I_{22} is negative, $\frac{d^2}{dz^2} F(z)$ is negative, so that $F(z)$ is an upward convex function of z.

Appendix A.2

Let

$$F(\beta_1,\beta_2) \equiv \int_0^\infty \int_0^\infty e^{-x_1-x_2}\, x_1^{-\frac{1}{2}}\, x_2^{-\frac{1}{2}} \left\{ \int_0^T e^{-y}\, (y_1)^{\nu/2-1}\, dy \right\} dx_1 dz_2 \qquad \ldots \text{(A.2.1)}$$

where

$$T = (\beta_1 x_1 + \beta_2 x_2)/A', \quad A' = A/\nu, \quad \beta_1 + \beta_2 = 1 \text{ and } \beta_1, \beta_2 > 0,$$

$$\frac{d}{d\beta_1} F(\beta_1, \beta_2) = K_1 \int_0^\infty \int_0^\infty e^{-x_1(1+\beta_1/A')-x_2(1+\beta_2/A')}\, x_1^{-\frac{1}{2}}\, x_2^{-\frac{1}{2}} \{\beta_1 x_1 + \beta_2 x_2\}^{\nu/2-1} (x_1-x_2) dx_1 dx_2$$

$$\qquad \ldots \text{(A.2.2)}$$

$$= K_2 \int_0^\infty \int_0^\infty e^{-u_1-u_2}\, u_1^{-\frac{1}{2}} u_2^{-\frac{1}{2}} \{c_1 u_1 + c_2 u_2\}^{\nu/2-1} \times \{u_1(1+a_1)^{-1} - u_2(1+a_2)^{-1}\} du_1 du_2$$

$$\qquad \ldots \text{(A.2.3)}$$

where

$$a_1 = \beta_1/A'; \quad a_2 = \beta_2/A';$$
$$c_1 = \beta_1/(1+a_1); \quad c_2 = \beta_2/(1+a_2).$$

Sub-case 1 : For $\nu = 2$, from (A.2.3.)

$$\frac{d}{d\beta_1} F(\beta_1, \beta_2) = I_1 - I_2. \qquad \ldots \text{(A.2.4)}$$

where

$$I_1 = (1+a_1)^{-1} \int_0^\infty \int_0^\infty e^{-u_1-u_2}\, u_1^{-\frac{1}{2}}\, {}_2^{-\frac{1}{2}}\, u_1 du_1 du_2$$

$$I_2 = (1+a_2)^{-1} \int_0^\infty \int_0^\infty e^{-u_1-u_2}\, u_1^{-\frac{1}{2}}\, u_2^{-\frac{1}{2}}\, u_2 du_1 du_2$$

From (A.2.4)
$$I_2/I_1 = (1+a_1)/(1+a_2) = (A'+\beta_1)/(A'+\beta_2)$$

which is less than unity if $\beta_1 < \beta_2$, so that $F(\beta_1, \beta_2)$ increases as β_1 increases $(\beta_1 < \beta_2)$. It can be similarly shown that if $\beta_1 > \beta_2$ $F(\beta_1, \beta_2)$ decreases as β_1 increases and the function $F(\beta_1, \beta_2)$ has a maximum value at $\beta_1 = \beta_2 = 1/2$.

Sub-case 2 : For $\nu = 1$, from (A.2.3) for $\beta_1, \beta_2 \geqslant \epsilon > 0$,

$$\frac{d}{d\beta_1} F(\beta_1, \beta_2) = I_1 - I_2 \qquad \ldots \text{(A.2.5)}$$

where

$$I_1 = K_2 (1+a_1)^{-1} \int_0^\infty \int_0^\infty e^{-u_1-u_2}\, u_1^{-\frac{1}{2}}\, u_2^{-\frac{1}{2}} (c_1 u_1 + c_2 u_2)^{-\frac{1}{2}} u_1 du_1 du_2$$

and

$$I_2 = K_2 (1+a_2)^{-1} \int_0^\infty \int_0^\infty e^{-u_1-u_2}\, u_1^{-\frac{1}{2}}\, u_2^{-\frac{1}{2}} (c_1 u_1 + c_2 u_2)^{-\frac{1}{2}} u_2 du_1 du_2.$$

Defining variates $V_1 = u_1$ and $V_2 = u_2/u_1$, it can be shown that

$$I_1 = K_3 (1+a_1)^{-1} \int_0^\infty V_2^{-\frac{1}{2}} \{1+c_2 V_2/c_1\}^{-\frac{1}{2}} (1+V_2)^{3/2}\, dV_2. \qquad \ldots \text{(A.2.6)}$$

For $\beta_1 < \beta_2$, defining $Z = 1/(1+V_2)$, it can be shown that

$$I_1 = K_4 (1+a_1)^{-1} F\{1/2, 3/2; 2; \lambda_1\} \qquad \ldots \text{(A.2.7)}$$

where $\lambda_1 = A'(A'+\beta_1)^{-1}(\beta_2-\beta_1)/\beta_2$.

Also, for $\beta_1 < \beta_2$ it can be shown that

$$I_2 = K_4 (1+a_2)^{-1} F\{1/2, 1/2; 2; \lambda_1\}. \qquad \ldots \text{(A.2.8)}$$

From (A.2.7) and (A.2.8) it thus follows that for $\beta_1 < \beta_2$

$$I_2/I_1 = (A'+\beta_1)(A'+\beta_2)^{-1}\, F\{1/2, 1/2; 2; \lambda_1\}/F\{1/2, 3/2; 2; \lambda_1\} \qquad \ldots \text{(A.2.9)}$$

For $\beta_1 < \beta_2$ thus $F(\beta_1, \beta_2)$ increases as β_1 increases. It can also be similarly shown that for $\beta_1 > \beta_2$, $F(\beta_1, \beta_2)$ decreases as β_1 increases and the function has a maximum value at $\beta_1 = \beta_2 = 1/2$.

Sub-case 3 : For $\nu > 3$, we have from (A.2.3)

$$\frac{d}{d\beta_1} F(\beta_1, \beta_2) = I_1 - I_2 \qquad \dots \text{ (A.2.10)}$$

where

$$I_1 = K_2(1+a_1)^{-1} \int_0^\infty \int_0^\infty e^{-u_1 - u_2} u_1^{-\frac{1}{2}} u_2^{-\frac{1}{2}} \times \{c_1 u_1 + c_2 u_2\}^{\nu/2-1} u_1 du_1 du_2 \qquad \dots \text{ (A.2.11)}$$

and

$$I_2 = K_2(1+a_2)^{-1} \int_0^\infty \int_0^\infty e^{-u_1 - u_2} u_1^{-\frac{1}{2}} u_2^{-\frac{1}{2}} \times \{c_1 u_1 + c_2 u_2\}^{\nu/2-1} u_2 du_1 du_2. \qquad \dots \text{ (A.2.12)}$$

For $\beta = 0$, from (A.2.11) and (A.2.12)

$$I_1 = K_3. \Gamma(3/2)\ \Gamma(\nu/2 - \tfrac{1}{2})/(1+a_1),$$
$$I_2 = K_3. \Gamma(\tfrac{1}{2})\ \Gamma(\nu/2 + 1 - \tfrac{1}{2})/(1+a_2).$$

so that

$$I_2/I_1 = (1+a_1)(1+a_2)^{-1}(\nu-1) = (A'+\beta_1)(A'+\beta_2)^{-1}(\nu-1). \qquad \dots \text{ (A.2.13)}$$

From (A.2.13), I_2 would be greater than I_1 if

$$A'\nu = A > \nu/(\nu-2), \qquad \dots \text{ (A.2.14)}$$

Now for $0 < \beta_1 < \beta_2$, defining variates $V_2 = u_2$ and $V_1 = u_1/u_2$, it can be shown from (A.2.11) that

$$I_1 = K_4 (1+a_1)^{-1} \int_0^\infty V_1^{\frac{1}{2}} (1+D_1 V_1)^p (1+V_1)^{-(p+2)}\ dV_1 \qquad \dots \text{ (A.2.15)}$$

where

$$D_1 = c_1/c_2 \text{ and } p = \nu/2 - 1.$$

From (A.2.15) it can be shown that

$$I_1 = K_5 (1+a_1)^{-1} D_1^{-\frac{1}{2}} F\{\nu/2+1,\ \tfrac{1}{2}\ ;\ 2\ ;\ \lambda_2\} \qquad \dots \text{ (A.2.16)}$$

where

$$\lambda_2 = 1 - D_1 = 1 - \beta_1 \beta_2^{-1} (A'+\beta_2)/(A'+\beta_1).$$

It can also be shown from (A.2.12) that for $0 < \beta_1 < \beta_2$

$$I_2 = K_5 (1+a_2)^{-1} D_1^{\frac{1}{2}} F\{\nu/2+1, 3/2\ ;\ 2\ ;\ \lambda_2\}. \qquad \dots \text{ (A.2.17)}$$

From (A.2.16) and (A.2.17) we thus have

$$I_2/I_1 = D_1 (A'+\beta_1)(A'+\beta_2)^{-1} F\{\nu/2+1, 3/2\ ;\ 2\ ;\ \lambda_2\}/F\{\nu/2+1, 1/2\ ;\ 2\ ;\ \lambda_2\}$$
$$= (1-\lambda_2)(A'+\beta_1)(A'+\beta_2)^{-1} F\{\nu/2+1, 3/2\ ;\ 2\ ;\ \lambda_2\}/F\{\nu/2+1, 1/2\ ;\ 2\ ;\ \lambda_2\} \quad \dots \text{ (A.2.18)}$$

Now according to algebraic relations due to Gauss (Erdelyi, 1953) satisfied by contiguous hypergeometric functions,

$$(1-z)\ \frac{F(a, b+1\ ;\ c\ ;\ z)}{F(a, b\ ;\ c\ ;\ z)} = 1 + z\ \frac{a-c}{c}\ \frac{F(a, b+1\ ;\ c+1\ ;\ z)}{F(a, b\ ;\ c\ ;\ z)}. \qquad \dots \text{ (A.2.19)}$$

From (A.2.18) and (A.2.19) I_2 would be greater than I_1 if

$$(A'+\beta_1)(A'+\beta_2)^{-1} \{1 + \lambda_2 E\} > 1 \qquad \dots \text{ (A.2.20)}$$

where

$$E_1 = (a-c)\ c^{-1}\ F(a, b+1\ ;\ c+1\ ;\ \lambda_2)/F(a, b\ ;\ c\ ;\ \lambda_2)$$
$$a = \nu/2+1\ ;\ b = \tfrac{1}{2}\ ;\ c = 2\ ;$$

and

$$\lambda_2 = 1 - \beta_1.\beta_2^{-1} (A'+\beta_2)/(A'+\beta_1)^{-1}.$$

From (A.2.20), I_2 would be greater than I_1 if

$$(A'+\beta_1).(A'+\beta_2)^{-1} (1+E) > 1 + \beta_1\ \beta_2^{-1}\ E,$$

or, if

$$A'(\beta_2 - \beta_1)E > \beta_2(\beta_2 - \beta_1),$$

or, if

$$A \nu^{-1} E > \beta_2 \qquad \dots \text{ (A.2.21)}$$

Now λ_2 and β_2 are connected as

$$\lambda_2 = 1 - \beta_1 \cdot \beta_2^{-1} (A' + \beta_2)(A' + \beta_1)^{-1} = A' \beta_2^{-1} (\beta_2 - \beta_1)(A' + \beta_1)^{-1} \text{ so that for } 1 > \beta_2 \geqslant \varepsilon > \tfrac{1}{2},$$

$$\lambda_2 \geqslant A' \{A' + (1-\varepsilon)\}^{-1} \cdot (2\varepsilon - 1)\varepsilon^{-1} . \qquad \qquad \dots \text{(A.2.22)}$$

For clarity of exposition (A.2.21) would be considered under two heads :

Sub-case 1 : β_2 lies in the range $3/4 > \beta_2 > \dfrac{1}{2}$.

Sub-case 2 : β_2 lies in the range $1 > \beta_2 \geqslant 3/4$.

For sub-case 1, from (A.2.21) it follows that since $F(a, b+1; c+1; \lambda_2)/F(a, b; c; \lambda_2)$ is greater than or equal to unity, (A.2.21) would be satisfied if

$$A' = A/\nu > c (a-c)^{-1} \cdot 3/4 = 3/(\nu-2). \qquad \qquad \dots \text{(A.2.23)}$$

For sub-case 2, since λ_2 from (A.2.22) would be greater than or equal to $A'(A' + 1/4)^{-1} \cdot 2/3$, (A.2.21) would be satisfied if

$$A' \left[\frac{1 + a(b+1)\lambda_0/(c+1)}{1 + ab\lambda_0/c} \right] > c/(a-c) = 4/(\nu-2) \qquad \qquad \dots \text{(A.2.24)}$$

where

$$\lambda_0 = A'(A' + 1/4)^{-1} \cdot 2/3.$$

From (A.2.24) it follows that I_2 would be greater that I_1 if

$$A' \frac{A' + 1/4 + aA'/3}{A' + 1/4 + aA'/6} > 4/(\nu-2). \qquad \qquad \dots \text{(A.2.25)}$$

Considering (A.2.25) the following auxiliary function U may be considered :

$$U = A'\{(A' + 1/4) + aA'/3\} - 4(\nu-2)^{-1}\{A' + 1/4 + aA'/6\}. \qquad \dots \text{A.2.26}$$

In (A.2.26) substituting $K/(\nu-2)$ for A' we get

$$(\nu-2) U = (A' + 1/4)(K-4) + aA'(2K-4)/6$$

$$= A'\{K-4 + (\nu+2)(K-2)/6\} + (K-4)/4$$

or,

$$12(\nu-2)^2 U = K\{12(K-4) + 2(\nu+2)(K-2)\} + 3(K-4)(\nu-2)$$

$$= K^2(2\nu+16) - K(\nu+62) - 12(\nu-2). \qquad \qquad \dots \text{(A.2.27)}$$

Since the co-efficient of K^2 of the quadratic on the RHS of (A.2.27) is positive, for some value of $K > K_0$ numerical value of the quadratic and as such numerical value of U is positive. Let the roots of the quadratic

$$(2\nu+16)K^2 - K(\nu+62) - 12(\nu-2) = 0 \qquad \qquad \dots \text{(A.2.28)}$$

be K_1 and K_2 (where $K_2 > K_1$). Now it can be shown that for $\nu \geqslant 3$,

$$K_2 < \frac{(\nu+62) + (10\nu+43)}{4\nu+32} = \frac{11\nu+105}{4\nu+32} . \qquad \qquad \dots \text{(A.2.29)}$$

Since the expression on the RHS of (A.2.29) for $\nu \geqslant 3$ is less than $16/5$, it follows that U would be positive for $K \geqslant 16/5$, which means that (A.2.25) or (A.2.24) would be satisfied for

$$A'(= A/\nu) > 3.2/(\nu-2)$$

or,

$$A > 3.2\nu/(\nu-2). \qquad \qquad \dots \text{(A.2.30)}$$

From (A.2.14), (A.2.23) and (A.2.25) it thus follows that for $0 \leqslant \beta_1 < \beta_2$, I_2 would be greater than I_1 if

$$A \geqslant 3.2\nu/(\nu-2). \qquad \qquad \text{(A.2.31)}$$

The function $F(\beta_1, \beta_2)$ thus decreases as β_1 increases for $\beta_1 < \beta_2$, if A is greater than or equal to $3.2\nu/(\nu-2)$. It can also be similarly shown that for $\beta_1 > \beta_2$, $F(\beta_1, \beta_2)$ increases as β_1 increases if A is numerically greater than or equal to $3.2\nu/(\nu-2)$ and the function has a minimum value at $\beta_1 = \beta_2 = 1/2$ and maximum value at $\beta_1 = 0$ and $\beta_1 = 1$.

Since critical values of F-table for 1 and $v(v \geqslant 3)$ d.f. for 5 p.c., 2 p.c., 1 p.c. etc. level of significance are all greater than $3.2v/(v-2)$ [a relation which can be proved using the algebraic relation due to Fisher (1941, page 151 middle)] the relation :

$$\beta_1 \chi_{11}^2 + \beta_2 \chi_{12}^4 > 4\chi_{\nu}^2/\nu$$

would be satisfied with probability less than or equal to α for $\alpha = 0.05, 0.02, 0.01$, etc. and $v \geqslant 3$, if A is taken from F-table corresponding to 1 and v d.f. for given α.

REFERENCES

BANERJEE, S. K. (1960) : Approximate confidence interval for linear functions of means of K-population when the population variances are not equal. *Sankhyā*, **22**, 357-358.

————— (1961) : On confidence interval for two means problem based on separate estimates of variances and tabulated values of t-table. *Sankhyā*, Series A, **23**, 359-378.

BARNARD, M. M. (1935) : The secular variation of skull characters in four series of Egyptian skulls. *Ann. Eugen.*, **6**, 352-371.

COURANT, R. (1957) : *Differential and Integral Calculus*, Volume II, Blackie & Son Limited, London and Glasgow.

ERDELYI, A. (1953) : *Higher Transcendental Functions*, Volume 1, McGraw Hill Book Company, Inc.

FERRAR, W. L. (1953) : *Algebra*, Oxford University Press.

FISHER, R. A. (1941) : The asymptotic approach to Behrens integrals. *Ann. Eugen.*, **11**, 141-172.

A SEQUENTIAL TEST OF FIT FOR MULTIVARIATE DISTRIBUTIONS*

LIONEL WEISS

Cornell University

SUMMARY. $X_1, X_2, \ldots$ are independent and identically distributed k-dimensional random variables, with an unknown continuous distribution. The problem is to test the hypothesis that the distribution is a given completely specified distribution. A sequential test is carried out as follows. After $X_1, \ldots, X_n$ have been observed, a k-dimensional sphere is constructed around each X in such a manner that the set-theoretic sum of the n spheres is assigned a predetermined probability p_0 by the distribution being tested. The event B_n is defined as occurring if X_{n+1} falls in one of the n spheres. The events $B_1, B_2, \ldots$ are treated formally as though they are independent events with a common unknown probability p, and the Wald sequential test of the hypothesis that p is equal to p_0 against the alternative that p is equal to p_1 is applied. The original hypothesis is accepted if and only if the Wald test accepts the hypothesis that p is equal to p_0. The properties of the resulting test of fit are discussed.

1. DESCRIPTION OF THE TEST

$X_1, X_2, \ldots$ are independent, identically distributed k-dimensional random variables, with an unknown common probability density function. The hypothesis to be tested is that the unknown probability density function is $g(x)$, a completely specified density function. Here x denotes a k-dimensional vector, as it will throughout this paper. Let C denote the closed k-dimensional unit cube. It will be convenient to apply the transformation described by Rosenblatt (1952) to each of the variables $X_1, X_2, \ldots$, so that when the hypothesis is true, each transformed variable has the uniform distribution over C. This transformation also guarantees that each transformed variable will be in C whether the hypothesis is true or not. From now on we assume that this transformation has been applied to each of the variables $X_1, X_2, \ldots$, so that all probability density functions considered assign probability one to C, and we are testing the hypothesis that the common probability density function of $X_1, X_2, \ldots$ is uniform over C. The pre-assigned level of significance will be denoted by α.

$X_1, X_2, \ldots$ are observed sequentially. After $X_1, \ldots, X_n$ have been observed, we define a subset $S_n(t)$ of C as follows, for t positive. A point x of C is in $S_n(t)$ if and only if the closed k-dimensional sphere with center at x and with volume t/n contains at least one of the points $X_1, \ldots, X_n$.

We choose values p_0, p_1, and β, where $0 < p_0 < p_1 < 1$ and $0 < \beta < 1-\alpha$, and hold them fixed. We define $T(n)$ as the value of t for which the volume of the set $S_n(t)$ is exactly equal to p_0. Clearly, $T(n)$ is uniquely defined. We define the event B_n as occurring when and only when X_{n+1} is in $S_n(T(n))$.

The test of fit is carried out by acting as though $B_1, B_2, \ldots$ are independent events, each with the same unknown probability p, and using the Wald sequential test of the hypothesis that $p = p_0$ against the alternative that $p = p_1$, with level of

significance α and power $1-\beta$. To be more specific, for each positive integer m let D_m denote the number of events $B_1, B_2, \ldots, B_m$ which occur and define (Wald, 1947, p. 92)

$$a_m = \frac{\log [\beta/(1-\alpha)]}{\log (p_1/p_0) - \log [(1-p_1)/(1-p_0)]} + m \frac{\log [(1-p_0)/(1-p_1)]}{\log (p_1/p_0) - \log [(1-p_1)/(1-p_0)]}$$

$$r_m = \frac{\log [(1-\beta)/\alpha]}{\log (p_1/p_0) - \log [(1-p_1)/(1-p_0)]} + m \frac{\log[(1-p_0)/(1-p_1)]}{\log (p_1/p_0) - \log [(1-p_1)/(1-p_0)]} .$$

Sampling continues as long as $a_m < D_m < r_m$. The first time that these inequalities do not hold, we accept the hypothesis of uniform distribution if $D_m \leqslant a_m$, and reject the hypothesis if $D_m \geqslant r_m$.

This test of fit is an extension of a sequential test for the case $k = 1$ which has been discussed by the author (Weiss, 1961). In Section 3 we discuss the properties of the test, and in the next section we develop a basic theorem which will be used to develop the properties of the test.

2. A BASIC THEOREM

In this section we state and prove a theorem that will be basic for our investigation of the properties of the test.

A probability density function $f(x)$ will be said to have the property R if $f(x)$ assigns probability 1 to C, is bounded over C, and if for each point x in the interior of C the probability assigned by $f(x)$ to a sphere centered at x and of volume v and entirely contained in C can be written as $f(x)v + R(x, v)v^{1+1/k}$, where $|R(x, v)| < K_1$ for all x in C and for all $v < K_2$, where K_1 and K_2 are some finite constants. We note that the uniform density function has the property R, as does any bounded density function with bounded continuous first partial derivatives in the interior of C. The symbols dx and dy will be understood to denote $dx_1 \ldots dx_k$ and $dy_1 \ldots dy_k$ respectively. All regions of integration considered will be subsets of C; in particular, if an indicated region of integration is not a subset of C, the actual region of integration is to be understood to be the intersection of the indicated region with C. For given functions $h(x), f(x)$, and a given positive value t, the integral $\int_C h(x) \exp [-tf(x)]dx$ will be denoted by $A(t; f, h)$.

The following theorem is basic.

Theorem : *If the common probability density function of $X_1, X_2, \ldots$ is $f(x)$, where $f(x)$ has the property R, and $h(x)$ is any bounded probability density function over C, then $\int_{S_n(t)} h(x)dx$ converges to $1 - A(t; f, h)$ with probability 1 as n increases, for any given positive t.*

Proof : Throughout the proof, the quantities $\theta_1, \theta_2, \ldots$ will denote finite positive constants, whose actual values will not have to be specified.

Fix a positive value of t. Break C into two mutually exclusive parts: C_n' consists of those points of C whose distance to the nearest boundary of C is greater than $\left[\frac{\Gamma(\frac{1}{2}k+1)}{\Pi^{k/2}} \frac{t}{n} \right]^{1/k}$; C_n° consists of all other points of C. We note that if x is any

point in C_n', the sphere of volume t/n centered at x is entirely contained in C. The volume of C_n^* is $1 - \left[1 - \left(\frac{2\Gamma(\frac{1}{2}k+1)}{\Pi^{k/2}} \frac{t}{n}\right)^{1/k}\right]^k$, which is less than $\theta_1(t/n)^{1/k}$.

Denote by $F(x, t/n)$ the probability assigned by $f(x)$ to a sphere centered at x and of volume t/n. Since $f(x)$ has the property R, we have $F(x, t/n) = f(x)t/n + R(x, t/n)(t/n)^{1+1/k}$ for each x in C_n', where $|R(x, t/n)| < \theta_3$ for all x in C_n' and all sufficiently large n.

For each x in C, we define $I_n(x)$ to be equal to zero if at least one of $X_1, ..., X_n$ falls in the sphere of volume t/n centered at x, and define $I_n(x)$ to be equal to one if none of $X_1, ..., X_n$ falls in the sphere of volume t/n centered at x. Then

$$1 - \int_{S_n(t)} h(x)dx = \int_C h(x)I_n(x)dx. \qquad \text{...} \quad (2.1)$$

From Kolmogoroff [(1946), pp. 39–41 (see also Robbins (1944)],

we have
$$E\left\{\int_C h(x)I_n(x)dx\right\} = \int_C h(x)E\{I_n(x)\}dx$$
$$= \int_C h(x)[1 - F(x, t/n)]^n \, dx. \qquad \text{...} \quad (2.2)$$

We can write

$$\int_C h(x)[1 - F(x, t/n)]^n \, dx = \int_{C_n'} h(x)[1 - f(x)t/n - R(x, t/n)(t/n)^{1+1/k}]^n \, dx$$
$$+ \int_{C_n^*} h(x)[1 - F(x, t/n)]^n \, dx. \qquad \text{...} \quad (2.3)$$

Clearly, since $h(x)$ is bounded, the second integral on the right side of (2.3) is non-negative and less than $\theta_4(t/n)^{1/k}$. We investigate the first integral on the right side of equation (2.3). We have

$$\log[1 - f(x)t/n - R(x, t/n)(t/n)^{1+1/k}]^n = n \log[1 - f(x)t/n - R(x, t/n)(t/n)^{1+1/k}]$$
$$= -t f(x) + R_1(x, n)/n^{1/k} \qquad \text{...} \quad (2.4)$$

where $|R_1(x, n)| < \theta_5$ for all x in C_n' and all sufficiently large n. Therefore, the first integral on the right side of equation (2.3) can be written as

$$\int_{C_n'} h(x) \exp[-t f(x) + R_1(x, n)/n^{1/k}]dx$$
$$= \int_{C_n'} h(x)[1 + R_2(x, n)/n^{1/k}] \exp[-t f(x)]dx \qquad \text{...} \quad (2.5)$$

where $|R_2(x, n)| < \theta_6$ for all x in C_n' and all sufficiently large n. From (2.5) it follows that the first integral on the right side of equation (2.3) can be written as

$$\int_{C_n'} h(x) \exp[-t f(x)]dx + R_3(n)/n^{1/k} \qquad \text{...} \quad (2.6)$$

where $|R_3(n)| < \theta_7$ for all sufficiently large n. Also, because of the bound on the volume of C_n^* given above, the integral in (2.6) differs from $A(t; f, h)$ by less than $\theta_8/n^{1/k}$. Then we have from (2.1)–(2.6)

$$E[1 - \int_{S_n(t)} h(x)dx] = A(t; f, h) + R_4(n)/n^{1/k} \qquad \text{...} \quad (2.7)$$

where $|R_4(n)| < \theta_9$ for all sufficiently large n.

Next we investigate

$$E\{[1- \int_{S_n(t)} h(x)dx - A(t; f, h)]^2\} \qquad \ldots \text{(2.8)}$$

which can be written as

$$E\{[\int_C h(x)I_n(x)dx]^2\} - 2A(t; f, h)\, E\{\int_C h(x)\, I_n(x)dx\} + A^2(t; f, h). \qquad \ldots \text{(2.9)}$$

The first term of (2.9) can be written as

$$E\{\int_C \int_C h(x)\, h(y)I_n(x)I_n(y)\, dy\, dx\}. \qquad \ldots \text{(2.10)}$$

For our investigation of (2.10) we introduce the following notation. $F(x, y, t/n)$ denotes the probability assigned by $f(x)$ to the set-theoretic sum of the spheres of volume t/n centered at x, y respectively. Then

$$E\{I_n(x)I_n(y)\} = [1-F(x, y, t/n)]^n. \qquad \ldots \text{(2.11)}$$

For a given point x in C'_n, we denote by $C_n(x)$ the set of all points y which are in C'_n and which are farther from x than $2\left[\dfrac{\Gamma(\frac{1}{2}k+1)}{\Pi^{k/2}}\dfrac{t}{n}\right]^{1/k}$. We note that if x is in C'_n and y is in $C_n(x)$, the spheres of volume t/n centered at x, y respectively are disjoint and are both contained in C, and therefore

$$[1-F(x, y, t/n)]^n = \left[\begin{array}{c} 1-f(x)t/n-f(y)t/n-R(x, t/n)(t/n)^{1+1/k} \\ -R(y, t/n)(t/n)^{1+1/k} \end{array}\right]^n. \qquad \ldots \text{(2.12)}$$

Denote by $C_n^*(x)$ the set of all points y which are in C but not in $C_n(x)$. The volume of $C_n^*(x)$ is no greater than the sum of the volume of C_n^* and the volume of a sphere of radius $2\left[\dfrac{\Gamma(\frac{1}{2}k+1)}{\Pi^{k/2}}\dfrac{t}{n}\right]^{1/k}$, and it follows easily that the volume of $C_n^*(x)$ is less than $\theta_{10}/n^{1/k}$.

We can write (2.10) as

$$\int_{C'_n}\{\int_{C_n(x)} h(x)\, h(y)[1-F(x, y, t/n)]^n dy\}dx$$

$$+\int_{C'_n}\{\int_{C_n^*(x)} h(x)\, h(y)[1-F(x, y, t/n)]^n dy\}dx$$

$$+\int_{C_n^*}\{\int_C h(x)\, h(y)[1-F(x, y, t/n)]^n dy\}dx \qquad \ldots \text{(2.13)}$$

and from (2.12) and the known bounds on $h(x)$ and the volumes of C_n^* and $C_n^*(x)$, (2.13) can be written as

$$\int_{C'_n}\int_{C_n(x)} h(x)\, h(y)\left[\begin{array}{c} 1-\{f(x)+f(y)\}\dfrac{t}{n} \\ -\Big\{R\Big(x, \dfrac{t}{n}\Big)-R\Big(y, \dfrac{t}{n}\Big)\Big\}\Big(\dfrac{t}{n}\Big)^{1+1/k} \\ +R_5(n)/n^{1/k} \end{array}\right]^n dy\, dx \qquad \ldots \text{(2.14)}$$

where $|R_5(n)| < \theta_{11}$ for all sufficiently large n. Just as in the discussion following (2.4), we find that the integral appearing in (2.14) can be written as

$$\int_{C'_n}\int_{C_n(x)} h(x)\, h(y)\exp[-t f(x)-t f(y)]\, dy\, dx + R_6(n)/n^{1/k} \qquad \ldots \text{(2.15)}$$

where $|R_6(n)| < \theta_{12}$ for all sufficiently large n. From the bounds on the volumes of C_n^* and $C_n^*(x)$, the integral appearing in (2.15) can be written as

$$\int_C \int_C h(x)\, h(y)\exp[-t f(x)-t f(y)]dy\, dx + R_7(n)/n^{1/k}$$

$$= A^2(t; f, h) + R_7(n)/n^{1/k} \qquad \ldots \text{(2.16)}$$

where $|R_7(n)| < \theta_{13}$ for all sufficiently large n. Using (2.13)—(2.16), we find that (2.10) can be written as

$$A^2(t; f, h) + R_8(n)/n^{1/k} \qquad \dots \ (2.17)$$

where $|R_8(n)| < \theta_{14}$ for all sufficiently large n. Applying (2.17) and (2.7) to (2.9), we find that (2.8) can be written as

$$R_9(n)/n^{1/k} \qquad \dots \ (2.18)$$

where $|R_9(n)| < \theta_{15}$ for all sufficiently large n. Using (2.18) and Tchebysheff's inequality, we find that for any positive ϵ,

$$P[|\int_{S_n(t)} h(x)\,dx - [1 - A(t; f, h)]| \geqslant \epsilon] \leqslant \frac{\theta_{15}}{\epsilon^2\, n^{1/k}} \cdot \qquad \dots \ (2.19)$$

For a given positive ϵ and positive integer m, let $j(m; \epsilon)$ denote the event $|\int_{S_m(t)} h(x)\,dx - [1 - A(t; f, h)]| < \epsilon$, and let $J(m; \epsilon)$ denote the simultaneous occurrence of all the events $j(m; \epsilon)$, $j(m+1; \epsilon)$, ... Let $K(m; \epsilon)$ denote the simultaneous occurrence of all the events $j(m^{2k}; \epsilon)$, $j((m+1)^{2k}; \epsilon)$, ... From (2.19) and elementary considerations, we have

$$P[K(m; \epsilon)] \geqslant 1 - \sum_{i=m}^{\infty} \frac{\theta_{15}}{\epsilon^2\, i^2}. \qquad \dots \ (2.20)$$

From (2.20), it follows that given any positive values ϵ, δ, there is a finite positive integer $M(\epsilon, \delta)$ such that if $m \geqslant M(\epsilon, \delta)$, then $P[K(m; \epsilon)] \geqslant 1 - \delta$.

Let θ_{16} be an upper bound on $h(x)$. If u, v are positive integers with $u \leqslant v$, then

$$\int_{S_v(t)} h(x)\,dx \leqslant \int_{S_u(t)} h(x)\,dx + \theta_{16}(v-u)t/v \qquad (2.21)$$

since the greatest volume each of the points $X_{u+1}, ..., X_v$ could contribute to $S_v(t)$ is t/v. Choose a positive integer w, and let n be an integer satisfying $w^{2k} < n \leqslant (w+1)^{2k}$. Using (2.21), we find

$$\int_{S_{(w+1)^{2k}}(t)} h(x)\,dx - \theta_{16}[(w+1)^{2k} - n]t/(w+1)^{2k} \qquad \dots \ (2.22)$$

$$\leqslant \int_{S_n(t)} h(x)\,dx \leqslant \int_{S_{w^{2k}}(t)} h(x)\,dx + \theta_{16}[n - w^{2k}]t/n.$$

Since $[(w+1)^{2k} - n]t/(w+1)^{2k}$ is no greater than $t\left[1 - \left(\frac{w}{w+1}\right)^{2k}\right]$ and $[n - w^{2k}]\, t/n$ is no greater than $t\left[\left(\frac{w+1}{w}\right)^{2k} - 1\right]$, it follows from (2.22) that given any ϵ, there is a finite positive integer $L(\epsilon)$ such that if $m \geqslant L(\epsilon)$, the occurrence of $K(m; \epsilon/2)$ implies the occurrence of $J(m^{2k}; \epsilon)$. Define $N(\epsilon, \delta)$ as the larger of the integers $L(\epsilon)$, $M(\epsilon/2, \delta)$. If $m \geqslant N(\epsilon, \delta)$, then $P[J(m^{2k}; \epsilon)] \geqslant 1 - \delta$. This completes the proof of the theorem.

Before applying the theorem to the sequential test of fit, we note that the theorem can be proved under less restrictive conditions, although the proof becomes more complicated. For example, if C can be broken into a finite number of regions, in each separate region $f(x)$ has the property R, and the set of all the boundary points of the regions has measure zero, the theorem holds.

It is easily seen that the convergence of $\int_{S_n(t)} h(x)\,dx$ to $1 - A(t; f, h)$ is uniform in t.

3. Properties of the sequential test

First we discuss the properties of the test when the hypothesis is true, that is, when the common distribution of X_1, X_2, ... is uniform over C. In this case, B_1, B_2, ... actually are independent events, with common probability p_0 since $P(B_n) = P[X_{n+1} \text{ in } S_n(T(n))] = \text{volume of } S_n(T(n)) = p_0$, and therefore the properties of the test of fit are exactly the same as the properties of the test on p when $p = p_0$: that is, the level of significance is approximately α, and the expected sample size is, from Wald [(1947), p. 100], approximately

$$\frac{(1-\alpha) \log [\beta/(1-\alpha)]+\alpha \log [(1-\beta)/\alpha]}{p_0 \log (p_1/p_0)+(1-p_0) \log [(1-p_1)/(1-p_0)]}.$$

(The word "approximately" is used because Wald obtained the properties of his sequential test only approximately, by ignoring the excess of the probability ratio over the decision boundaries. The formula for the expected sample size should really be increased by one, since the event B_n is defined using X_{n+1}, but this is a minor refinement).

Next we investigate the properties of the sequential test of fit when the hypothesis is not true. Throughout this part of the discussion, we assume that the common probability density function of X_1, X_2, ... is $f(x)$, where $f(x)$ has the property R. Define $t(f)$ as the solution in t of the equation

$$1- \int_C \exp [-t f(x)]dx = p_0 \qquad \qquad \text{... (3.1)}$$

and denote the quantity $\quad 1- \int_C f(x) \exp [-t(f) f(x)]dx \qquad \qquad \text{... (3.2)}$

by $Q(f)$. We will show that $P(B_n|X_1, ..., X_n)$ converges to $Q(f)$ with probability 1 as n increases. To show this, we set $h(x) = 1$ in the theorem of Section 2, and note that then the theorem states that $\int_{S_n(t)} 1 \, dx = \text{volume of } S_n(t)$ converges to $1-\int_C \exp [-t f(x)]dx$ with probability 1 as n increases. By the definitions of $T(n)$ and $t(f)$, this implies that $T(n)$ converges to $t(f)$ with probability one as n increases. Next we set $h(x) = f(x)$ in the theorem of Section 2, and note that then the theorem states that $P[X_{n+1} \text{ in } S_n(t)] = \int_{S_n(t)} f(x)dx$ converges to $1-\int_C f(x) \exp [-t f(x)]dx$ with probability 1 as n increases. Since $P(B_n|X_1, ..., X_n) = P[X_{n+1} \text{ in } S_n(T(n))]$, and we have seen that $T(n)$ converges to $t(f)$ with probability 1 as n increases, it follows that $P(B_n|X_1, ..., X_n)$ converges to $1-\int_C f(x) \exp [-t f(x)]dx = Q(f)$ with probability 1 as n increases.

Now we can compute the approximate power of the sequential test of fit against $f(x)$, if α and β are both small. For when α and β are both small, the test will surely take many observations before coming to a decision. Even large differences between $P(B_n|X_1, ..., X_n)$ and $Q(f)$ for a relatively few small values of n cannot have much effect on the power of the test, if many observations are taken. Since the properties of the Wald sequential test of a binomial parameter p vary continuously with p, it follows that when α and β are small, and the common density function of

$X_1, X_2, \ldots$ is $f(x)$, the power of the sequential test of fit is approximately the same as the power of the Wald test of p when p is actually equal to $Q(f)$. That is, from p. 96 of Wald (1947), if the quantity h is defined by the equation

$$Q(f) = \frac{1-[(1-p_1)/(1-p_0)]^h}{(p_1/p_0)^h - [(1-p_1)/(1-p_0)]^h}$$

then the probability that the hypothesis is rejected when the common density function is $f(x)$ is approximately

$$\frac{1-[\beta/(1-\alpha)]^h}{[(1-\beta)/\alpha]^h - [\beta/(1-\alpha)]^h}.$$

We note that a necessary condition for a reasonable power function for the sequential test is that $Q(f) > p_0$ for each density function $f(x)$ over C which differs from 1 on a subset of C of positive Lebesgue measure. It is clear that this inequality will be proved if we show that for each positive value of t,

$$1 - \int_C f(x) \exp[-t f(x)]\, dx \geqslant 1 - \int_C \exp[-t f(x)]\, dx$$

or equivalently, that

$$\int_C [1 - f(x)] \exp[-t f(x)]\, dx \geqslant 0 \qquad \ldots \text{(3.3)}$$

with equality if and only if $f(x) = 1$ almost everywhere on C. The integral in (3.3) can be written as

$$\int_{x:f(x)<1} [1-f(x)] \exp[-t f(x)] dx + \int_{x:f(x)>1} [1-f(x)] \exp[-t f(x)]\, dx. \qquad \ldots \text{(3.4)}$$

The first integral in (3.4) is greater than or equal to $\int_{x:f(x)<1} [1-f(x)]e^{-t}\, dx$, with equality if and only if the set $\{x : x \text{ in } C \text{ and } f(x) < 1\}$ has Lebesgue measure zero; the second integral in (3.4) is greater than or equal to $\int_{x:f(x)>1} [1-f(x)]e^{-t}\, dx$, with equality if and only if the set $\{x : x \text{ in } C \text{ and } f(x) > 1\}$ has Lebesgue measure zero. Therefore the integral in (3.3) is greater than or equal to

$$\int_{x:f(x)<1} [1-f(x)]e^{-t}\, dx + \int_{x:f(x)>1} [1-f(x)]e^{-t}\, dx = 0$$

with equality if and only if $f(x) = 1$ almost everywhere on C. This proves the desired inequality.

Another characteristic of the sequential test of interest is the expected sample size when the hypothesis of a uniform distribution is not true. It would seem a reasonable conjecture that when the common density function of $X_1, X_2, \ldots$ is $f(x)$, the expected sample size is approximately the expected sample size of the Wald test of a binomial parameter p when $p = Q(f)$. However, the sample size is an unbounded function over the sample sequences, and therefore even though the probability of large deviations between $P(B_n | X_1, \ldots, X_n)$ and $Q(f)$ are small, such deviations could still have a large disturbing effect on the expected sample size. More research must be done on this.

4. Comparison of the Sequential Test with Other Tests of Fit

There are only a few test of fit in the multivariate case which are generally known. The most familiar is the chi-square test which involves breaking C into subregions, and using the numbers of observations falling in the various subregions. Somewhat less familiar are the Kolmogoroff-Smirnov test and the von Mises test, both described by Rosenblatt (1952).

The different tests can be compared according to various criteria. For example, the purely mechanical problem of computing the test statistic can be formidable. In this respect, the chi-square test seems the most convenient, while the sequential test seems the least convenient.

Another criterion for comparing the tests is the amount of knowledge available to us about their power functions. We seem to know more about the power functions of the chi-square test and the sequential test than about the power functions of the two other tests. Also, our knowledge about the power function of the sequential test is in a very convenient form, in the following sense. Suppose we want a test of fit of level of significance α and power $1-\beta$ against a particular given alternative $f(x)$, where α and β are small. Then we simply set $p_1 = Q(f)$ in the specification of the sequential test to achieve this. Note that this is a reasonable way to choose a value for p_1, which up to now has been an arbitrary value in the open interval $(p_0, 1)$. This still leaves us the choice of the value of p_0. The lower the value chosen for p_0, the smaller is the expected sample size. This might seem desirable, but there is a danger here : if the expected sample size is made too small, the test will probably not continue long enough for the theorem of Section 2 to come into play, and we will lose our knowledge about the power of the test.

By far the most important criterion for the comparison of tests is the shape of their power functions. It seems unlikely that any one of the four tests mentioned will have a uniformly higher power function than one of the other tests. For example, it is easy to find alternatives against which the power of the chi-square test is low, by assigning to the various subregions into which C is broken the same probabilities under the alternative as under the hypothesis, while within the subregions the alternative and hypothesis can differ greatly.

Note : The test seems to be symmetric in the k components of X, in the sense that for a given sample sequence, consistently interchanging the roles of the coordinates of X will not affect the outcome of the test. This is so because the construction uses the Euclidean distance $\Sigma(y_i - z_i)^2$ between two points $(y_1, \ldots, y_k)$ and $(z_1, \ldots, z_k)$, and this distance is unaffected by interchanging, for example, y_1 with y_2 and z_1, with z_2. The test is not invariant under permutations of the observed vectors $X_1, X_2, \ldots$, but this is a familiar property of sequential tests.

References

Kolmogoroff, A. (1946) : *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Chelsea Publishing Company, New York.

Robbins, H. E. (1944) : On the measure of a random set. *Ann. Math. Stat.*, 15, 70–74.

Rosenblatt, M. (1952) : Remarks on a multivariate transformation. *Ann. Math. Stat.*, 23, 470–472.

Wald, A. (1947) : *Sequential Analysis*, John Wiley and Sons, New York.

Weiss, L. (1961) : Tests of fit based on the number of observations falling in the shortest sample spacings determined by earlier observations. *Ann. Math. Stat.*, 32, 838–845.

Paper received : December, 1961.

THE DISTRIBUTION OF THE RATIO OF THE VARIANCES
OF VARIATE DIFFERENCES IN THE CIRCULAR CASE*

By J. N. K. RAO and G. TINTNER
Iowa State University

SUMMARY. In time series analysis, the variate difference method is used to test the order of the finite difference at which the trend or the systematic part in the time series is approximately eliminated. There is no exact test available in the literature except for the one proposed by Tintner based on a method of selection which uses only a portion of the observations. In this paper, the statistic V_{k+1}/V_k is proposed to test that the trend is approximately eliminated at the k-th finite differencing of the series, where V_k is the variance of the series of the k-th differences. Its exact distribution assuming that the observations are $NI(0, \sigma^2)$ is derived under a circular definition of the universe. The lower 5% and 1% points of the statistics V_2/V_1 and V_3/V_2 are tabulated for various values of N, the size of the sample. In practice, one uses the non-circular statistic with these percentage points for the circular statistic as an approximation, especially with long time series.

1. INTRODUCTION

Statistics based on successive differences have been widely used for testing the independence of successive observations in the analysis of economic time series, etc. [e.g., Anderson (1929) and Tintner (1940)]. von Neumann (1942) proposed the statistic δ^2/s^2 for testing the independence of successive observations where δ^2 is the mean square successive first difference and s^2 is the variance of the observations. He obtained the distribution of δ^2/s^2 assuming that the observations are not auto-correlated. Percentage points of the distribution of this statistic have been obtained by Hart (1942). One disadvantageous feature of δ^2/s^2 is this: If there is a slow moving trend in the mean value, due to the appearance of the mean in s^2, it will be heavily biased. This difficulty could be avoided if the statistic V_2/V_1 is used, where V_2 is the variance of the series of the second differences and V_1 is the variance of the series of the first differences of the successive observations, since V_2 and V_1 are independent of the mean value and the bias in V_1 and V_2 is much smaller than the bias in s^2 if there is a trend in the mean (Kamat, 1954). Dixon (1944) derived the first two moments of V_2/V_1 approximately by smoothing the joint characteristic function of V_1 and V_2 using the method of Koopmans (1942), assuming a circular definition of the universe.

In time series analysis, the variate difference method is used to test the order of the finite difference at which the trend or the systematic part in the time series is approximately eliminated. There is no exact test available in the literature, except for the one proposed by Tintner (1940) based on a method of selection which uses only a portion of the observations. The statistic V_{k+1}^*/V_k^* where V_k^* and V_{k+1}^* are the variances of the k-th and $(k+1)$-th differences respectively, based upon selected observations can be used to test the above hypothesis. In this paper we derive the exact distribution of the ratio V_{k+1}/V_k under the circular definition assuming that the observations are not autocorrelated. The lower 5% and 1% points are tabulated for the statistic V_2/V_1 and are compared with those obtained by a normal approximation

using Dixon's (1944) first two moments of V_2/V_1. In practice, one can use the non-circular statistic with these percentage points for the circular statistic as an approximation, especially with long time series. Also, the lower 5% and 1% points of the statistic V_3/V_2 are tabulated which may be useful in employing the variate difference method.

2. The joint characteristic function of V_{k+1} and V_k

Let N successive observations in the sample be $x_1, ..., x_N$ and with the circular definition

$$x_{N+t} = x_{-N+t} = x_t. \qquad \text{... (2.1)}$$

The variance of the k-th difference is defined as

$$V_k = \frac{\sum\limits_{1}^{N} (\Delta^k x_t)^2}{N \binom{2k}{k}} \qquad \text{... (2.2)}$$

where $\Delta^k x_t$ is the k-th difference of the x_t. To find the joint characteristic function of V_k and V_{k+1} we use the method of Tintner (1955) in deriving the characteristic function of V_k. Under the null hypothesis, the observations $x_1, ..., x_N$ are assumed to be $NI(0, \sigma^2)$. Let $\phi(t_1, t_2)$ denote the joint characteristic function of V_{k+1} and V_k. Then

$$\phi(t_1, t_2) = \frac{1}{(2\pi)^{N/2}\sigma^N} \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} \cdots \int\limits_{-\infty}^{\infty} e^{i(t_1 V_{k+1} + t_2 V_k)} e^{-\Sigma x_t^2/2\sigma^2} dx_1\, dx_2 \dots dx_N \quad \text{... (2.3)}$$

where $i = \sqrt{-1}$. Expanding $\Sigma(\Delta^k x_t)^2$ and $\Sigma(\Delta^{k+1} x_t)^2$ we find

$$V_k = \frac{1}{N\binom{2k}{k}} \left[\binom{2k}{k}\Sigma x_t^2 - \binom{2k}{k-1}(\Sigma x_t x_{t+1} + \Sigma x_t x_{t-1}) + \dots \right.$$

$$\left. + (-1)^k \binom{2k}{0}(\Sigma x_t x_{t+k} + \Sigma x_t x_{t-k}) \right] \qquad \text{... (2.4)}$$

and

$$V_{k+1} = \frac{1}{N\binom{2k+2}{k+1}} \left[\binom{2k+2}{k+1}\Sigma x_t^2 - \binom{2k+2}{k}(\Sigma x_t x_{t+1} + \Sigma x_t x_{t-1}) + \dots \right.$$

$$+ (-1)^k \binom{2k+2}{1}(\Sigma x_t x_{t+k} + \Sigma x_t x_{t-k})$$

$$\left. + (-1)^{k+1} \binom{2k+2}{0}(\Sigma x_t x_{t+k+1} + \Sigma x_t x_{t-k-1}) \right]. \qquad \text{... (2.5)}$$

Using equations (2.4) and (2.5), $\phi(t_1, t_2)$ can be written as

$$\phi(t_1, t_2) = \frac{1}{(2\pi)^{N/2}\sigma^N} \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} \cdots \int\limits_{-\infty}^{\infty} e^{-x'Dx/2\sigma^2} dx_1\, dx_2 \dots dx_N \qquad \text{... (2.6)}$$

where x is the column vector of the x_i, and D is a matrix whose determinant $|D|$ is a circulant. Evaluation of $|D|$ gives

$$|D| = \prod_{j=0}^{N-1} [a_0 + a_1\omega^j + \ldots + a_{k+1}\omega^{(k+1)j} + a_1\omega^{-j} + \ldots + a_{k+1}\omega^{-(k+1)j}]. \quad \ldots \quad (2.7)$$

Here ω is the N-th root of unity,

$$\omega^j = \cos\frac{2\pi j}{N} + i\sin\frac{2\pi j}{N} \quad\quad \ldots \quad (2.8)$$

and

$$a_0 = 1 - \frac{2i\sigma^2 t_1}{N}\frac{\binom{2k+2}{k+1}}{\binom{2k+2}{k+1}} - \frac{2i\sigma^2 t_2}{N}\frac{\binom{2k}{k}}{\binom{2k}{k}}$$

$$a_1 = \frac{2i\sigma^2 t_1}{N}\frac{\binom{2k+2}{k}}{\binom{2k+2}{k+1}} + \frac{2i\sigma^2 t_2}{N}\frac{\binom{2k}{k-1}}{\binom{2k}{k}} \quad\quad \ldots \quad (2.9)$$

$$a_k = (-1)^{k+1}\frac{2i\sigma^2 t_1}{N}\frac{\binom{2k+2}{1}}{\binom{2k+2}{k+1}} + (-1)^{k+1}\frac{2i\sigma^2 t_2}{N}\frac{\binom{2k}{0}}{\binom{2k}{k}}$$

$$a_{k+1} = (-1)^{k+2}\frac{2i\sigma^2 t_1}{N}\frac{\binom{2k+2}{0}}{\binom{2k+2}{k+1}}.$$

Now using the relations

$$2^{2m}\left(\sin\frac{\pi j}{N}\right)^{2m} = \binom{2m}{m} - 2\binom{2m}{m-1}\cos\frac{2\pi j}{N} + \ldots + (-1)^m 2\binom{2m}{0}\cos\frac{2\pi m j}{N}$$

equation (2.7) can be simplified as

$$|D| = \prod_{j=0}^{N-1}\left[1 - \frac{2it_1\sigma^2}{N\binom{2k+2}{k+1}} 2^{2k+2}\left(\sin\frac{\pi j}{N}\right)^{2k+2} - \frac{2it_2\sigma^2}{N\binom{2k}{k}} 2^{2k}\left(\sin\frac{\pi j}{N}\right)^{2k}\right]. \quad \ldots \quad (2.10)$$

Now the evaluation of (2.6) gives

$$\phi(t_1, t_2) = |D|^{-\frac{1}{2}} = \prod_{j=0}^{N-1}\left[1 - \frac{2it_1\sigma^2}{N\binom{2k+2}{k+1}} 2^{2k+2}\left(\sin\frac{\pi j}{N}\right)^{2k+2}\right.$$

$$\left. - \frac{2it_2\sigma^2}{N\binom{2k}{k}} 2^{2k}\left(\sin\frac{\pi j}{N}\right)^{2k}\right]^{-\frac{1}{2}}. \quad \ldots \quad (2.11)$$

From (2.11) one can easily derive the following variances and correlation coefficients :

$$\sigma_{k+1} = \text{var}(V_{k+1}) = \frac{2\sigma^4 \binom{4k+4}{2k+2}}{N \binom{2k+2}{k+1}^2}$$

$$\sigma_k^2 = \text{var}(V_k) = \frac{2\sigma^4 \binom{4k}{2k}}{N \binom{2k}{k}^2} \qquad \qquad \dots \quad (2.12)$$

and
$$\rho_k^2 = \text{corr}^2(V_{k+1}, V_k) = \frac{\binom{4k+2}{2k+1}^2}{\binom{4k}{2k}\binom{4k+4}{2k+2}} = 1 - \frac{1}{8k^2+10k+3}. \qquad \dots \quad (2.13)$$

It is interesting to note from (2.13) that

$$\frac{20}{21} \leqslant \rho_k^2 \leqslant 1 \qquad \qquad \dots \quad (2.14)$$

and that ρ_k^2 increases monotonically with k.

Now we show that V_{k+1} and V_k have a joint bivariate normal distribution as $N \to \infty$. Consider the joint cumulant generating function of the standardized variates

$$V_{k+1}^1 = \frac{V_{k+1} - \sigma^2}{\sigma_{k+1}} \text{ and } V_k^1 = \frac{V_k - \sigma^2}{\sigma_k.} \qquad \qquad \cdot \quad \dots \quad (2.15)$$

which can be obtained from (2.11) as

$$h^1(t_1, t_2) = \log \phi^1(t_1, t_2) = - \frac{it_1\sigma^2}{\sigma_{k+1}} - \frac{it_2\sigma^2}{\sigma_k}$$

$$-\frac{1}{2} \sum_{j=0}^{N-1} \log \left[1 - 2it_1 \frac{2^{2k+2}\sigma^2}{\sigma_{k+1} N \binom{2k+2}{k+1}} \left(\sin \frac{\pi j}{N}\right)^{2k+2} - 2it_2 \frac{2^{2k}\sigma^2}{\sigma_k N \binom{2k}{k}} \left(\sin \frac{\pi j}{N}\right)^{2k} \right]$$

$$\dots \quad (2.16)$$

where $\phi^1(t_1, t_2)$ is the joint characteristic function of the standardized variates V_{k+1}^1 and V_k^1. Now expanding the logarithm in (2.16), using the relations

$$\sum_{j=0}^{N-1} 2^{2rm} \left(\sin \frac{\pi j}{N}\right)^{2rm} = N \binom{2rm}{rm} \qquad \qquad \dots \quad (2.17)$$

and making $N \to \infty$, after some simplification we find that

$$\phi^1(t_1, t_2) \sim e^{-\frac{1}{2}(t_1^2 + t_2^2 + 2\rho_k t_1 t_2)} \qquad \qquad \dots \quad (2.18)$$

which is the joint characteristic function of two variates with a bivariate normal distribution with zero means, unit variances and correlation coefficient ρ_k. Hence, as $N \to \infty$, V_{k+1} and V_k have a bivariate normal distribution with the same mean σ^2, variances σ_{k+1}^2 and σ_k^2 and correlation coefficient ρ_k.

3. The exact distribution of the ratio V_{k+1}/V_k

In order to find the exact distribution of the ratio V_{k+1}/V_k from the joint characteristic function $\phi(t_1, t_2)$, we use the inversion theorem for the ratio (Gurland, 1948). Now the cumulative density function of V_{k+1}/V_k is

$$G_k'(\lambda) = Pr\left(\frac{V_{k+1}}{V_k} \leqslant \lambda\right) = Pr(V_{k+1} - \lambda V_k \leqslant 0) \qquad \ldots \ (3.1)$$

and from Gurland (1948) it follows that

$$G_k(\lambda) = \tfrac{1}{2} - \frac{1}{2\pi i} \oint \frac{\phi(t, -t\lambda)}{t} \, dt \qquad \ldots \ (3.2)$$

where $\oint$ denotes the Cauchy principal value and $\phi(t, -t\lambda)$ is obtained by substituting $t_1 = t$ and $t_2 = -t\lambda$ in (2.11) for $\phi(t_1, t_2)$. Consider now the two cases N odd and N even. When N is odd, from (2.11) it is easily seen that

$$\phi(t, -t\lambda) = \prod_{j=1}^{(N-1)/2} (1 - a_j \, i \, t)^{-1} \qquad \ldots \ (3.3)$$

and when N is even $\quad \phi(t, -t\lambda) = (1 - c \, i \, t)^{-\frac{1}{2}} \prod_{j=1}^{(N-2)/2} (1 - a_j \, i \, t)^{-1} \qquad \ldots \ (3.4)$

where
$$a_j = 2\sigma^2 \frac{2^{2k+2}\left(\sin\dfrac{\pi j}{N}\right)^{2k+2}}{N\dbinom{2k+2}{k+1}} - 2\sigma^2\lambda \frac{2^{2k}\left(\sin\dfrac{\pi j}{N}\right)^{2k}}{N\dbinom{2k}{k}} \qquad \ldots \ (3.5)$$

and
$$c = 2\sigma^2 \frac{2^{2k+2}}{N\dbinom{2k+2}{k+1}} - 2\sigma^2\lambda \frac{2^{2k}}{N\dbinom{2k}{k}}. \qquad \ldots \ (3.6)$$

Case 1 : *N odd*. Now (3.3) can be put into partial fractions as

$$\phi(t, -t\lambda) = \sum_{j=1}^{(N-1)/2} \frac{B_j}{(1 - a_j \, i \, t)} \qquad \ldots \ (3.7)$$

where
$$B_j = \frac{a_j^{(N-3)/2}}{\prod_{r \neq s}^{(N-1)/2} (a_r - a_s)} \qquad \ldots \ (3.8)$$

Therefore, from (3.2) it follows that

$$G_k(\lambda) = \tfrac{1}{2} - \sum_{j=1}^{(N-1)/2} B_j \, F_1(j, \lambda) \qquad \ldots \ (3.9)$$

where
$$F_1(j, \lambda) = \frac{1}{2\pi i} \oint \frac{dt}{t(1 - a_j \, i \, t)}. \qquad \ldots \ (3.10)$$

By contour integration, the evaluation of (3.10) gives

$$F_1(j, \lambda) = \begin{cases} \tfrac{1}{2} & \text{if } a_j > 0 \\ -\tfrac{1}{2} & \text{if } a_j \leqslant 0 \end{cases} \qquad \text{... (3.11)}$$

Therefore

$$G_k(\lambda) = \tfrac{1}{2} - \tfrac{1}{2} \sum_{j=1}^{(N-1)/2} \delta_j B_j \qquad \text{... (3.12)}$$

where

$$\delta_j = \begin{cases} 1 & \text{if } a_j > 0 \\ -1 & \text{if } a_j \leqslant 0. \end{cases} \qquad \text{... (3.13)}$$

When $\lambda \leqslant 0$, $a_j > 0$ for all j so that

$$G_k(\lambda) = \tfrac{1}{2} - \tfrac{1}{2} \sum B_j = 0 \qquad \text{... (3.14)}$$

since $\sum B_j = 1$. Also for λ sufficiently large, say T, $a_j < 0$ for all j so that

$$G_k(T) = \tfrac{1}{2} + \tfrac{1}{2} \sum B_j = 1$$

and therefore

$$\lim_{T \to \infty} G_k(T) = G_k(\infty) = 1. \qquad \text{... (3.15)}$$

This provides a verification that $G_k(\lambda)$ given by (3.12) is in fact a cumulative density function.

Case 2 : N even. Now (3.4) can be put into partial fractions as

$$\phi(t, -t\lambda) = \sum_{j=1}^{(N-2)/2} \frac{A_j}{(1 - a_j i t)(1 - c i t)^{1/2}} \qquad \text{... (3.16)}$$

where

$$A_j = \frac{a_j^{(N-4)/2}}{\prod\limits_{r \neq s}^{(N-2)/2} (a_r - a_s)}. \qquad \text{... (3.17)}$$

Therefore, from (3.2) it follows that

$$G_k(\lambda) = \tfrac{1}{2} - \sum_{j=1}^{(N-2)/2} A_j F_2(j, \lambda) \qquad \text{... (3.18)}$$

where

$$F_2(j, \lambda) = \frac{1}{2\pi i} \oint \frac{dt}{t(1 - a_j i t)(1 - c i t)^{1/2}}. \qquad \text{... (3.19)}$$

Let $I_{m_1, m_2}(x)$ denote the distribution function of $\chi^2_{m_1} / \chi^2_{m_2}$, where $\chi^2_{m_1}$ and $\chi^2_{m_2}$ are χ^2 variables with m_1 and m_2 degrees of freedom respectively. Then we have from Gurland (1948)

$$I_{m_1, m_2}(x) = \tfrac{1}{2} - \frac{1}{2\pi i} \oint \frac{dt}{t(1 - 2 i t)^{m_1/2} (1 + 2 i t x)^{m_2/2}}. \qquad \text{... (3.20)}$$

To evaluate (3.19) we make use of (3.20) and distinguish the two cases $a_j > 0$ and $a_j < 0$.

The case $a_j > 0$. From (3.19) we have that

$$\tfrac{1}{2} - F_2(j, \lambda) = \tfrac{1}{2} - \frac{1}{2\pi i} \oint \frac{dt'}{t'(1 - 2it')\left(1 - \dfrac{2it'c}{a_j}\right)^{1/2}} \qquad \text{... (3.21)}$$

where $a_j t = 2t'$. Now comparing (3.21) and (3.20) it immediately follows that

$$\tfrac{1}{2} - F_2(j, \lambda) = I_{2,1}\left(-\frac{c}{a_j}\right). \qquad \text{... (3.22)}$$

It may be noted that c/a_j does not contain the nuisance parameter σ^2.

The case $a_j < 0$. In this case we have

$$\tfrac{1}{2} - F_2(j, \lambda) = \tfrac{1}{2} + \frac{1}{2\pi i} \oint \frac{dt'}{t'(1 - 2it')\left(1 - \dfrac{2it'c}{a_j}\right)^{1/2}} = 1 - I_{2,1}\left(-\frac{c}{a_j}\right). \text{... (3.23)}$$

Therefore

$$G_k(\lambda) = \sum_{j=1}^{(N-2)/2} \delta_j' A_j \qquad \text{... (3.24)}$$

where

$$\delta_j' = \begin{cases} I_{2,1}\left(-\dfrac{c}{a_j}\right) & \text{if } a_j > 0 \\[2mm] 1 - I_{2,1}\left(-\dfrac{c}{a_j}\right) & \text{if } a_j < 0. \end{cases} \qquad \text{... (3.25)}$$

The direct evaluation of $I_{2,1}\left(-\dfrac{c}{a_j}\right)$ gives

$$I_{2,1}\left(-\frac{c}{a_j}\right) = 1 - \left[\frac{1}{1 - (c/a_j)}\right]^{1/2}, \frac{c}{a_j} < 0. \qquad \text{... (3.26)}$$

Now when $\lambda \leqslant 0$, $a_j > 0$ and $c > 0$ so that $I_{2,1}\left(-\dfrac{c}{a_j}\right) = 0$

and

$$G_k(\lambda) = \tfrac{1}{2} - \tfrac{1}{2} \Sigma A_j = 0 \qquad \text{... (3.27)}$$

since $\Sigma A_j = 1$. When $\lambda = T$ where T is sufficiently large such that $a_j < 0$ and $c < 0$, then also $I_{2,1}\left(-\dfrac{c}{a_j}\right) = 0$ so that

$$G_k(T) = \tfrac{1}{2} + \tfrac{1}{2} \Sigma A_j = 1 \qquad \text{... (3.28)}$$

and hence

$$\underset{T \to \infty}{\mathrm{lt}}\; G_k(T) = G_k(\infty) = 1. \qquad \text{... (3.29)}$$

This provides a verification that $G_k(\lambda)$ given by (3.24) is in fact a cumulative density function.

4. Computation of Percentage Points for V_2/V_1 and V_3/V_2

The probabilities $G_1(\lambda)$ for several values of λ with an interval of 0.05 and for $N = 6$ to 25 are computed on IBM 650. Then using Bessel's second-degree inverse interpolation formula (since third differences are small), the values of λ corresponding to $G_1(\lambda) = 0.05$ and $G_1(\lambda) = 0.01$ denoted by $\lambda_{.05}^{(1)}$ and $\lambda_{.01}^{(1)}$ respectively, are obtained and are given in columns (2) and (5) of Table 1 below. The approximation of Dixon (1944) based on the smoothing process results in the following moments of V_2/V_1 :

$$E(V_2/V_1) = \frac{N+\frac{7}{8}}{N+1} \qquad \ldots \quad (4.1)$$

and

$$\text{var } (V_2/V_1) = \frac{(2N^2+7N+4)}{9(N+1)^2 \, (N+2)}. \qquad \ldots \quad (4.2)$$

TABLE 1. PERCENTAGE POINTS OF V_2/V_1 AND $\delta^2/2s^2$

| N | $\lambda_{.05}^{(1)}$ | | $\lambda_{.05}^{(0)}$ | $\lambda_{.01}^{(1)}$ | | $\lambda_{.01}^{(0)}$ |
| | exact | approximate | .05 | exact | approximate | .01 |
(1)	(2)	(3)	(4)	(5)	(6)	(7)
6	0.593	0.652	0.534	0.434	0.527	0.337
7	0.654	0.678	0.546	0.485	0.561	0.358
8	0.675	0.699	0.561	0.542	0.589	0.379
9	0.688	0.717	0.576	0.569	0.613	0.399
10	0.708	0.731	0.590	0.584	0.633	0.418
11	0.723	0.744	0.603	0.606	0.650	0.435
12	0.735	0.756	0.615	0.626	0.665	0.452
13	0.747	0.765	0.626	0.640	0.678	0.467
14	0.758	0.775	0.636	0.654	0.691	0.481
15	0.766	0.782	0.646	0.667	0.701	0.494
16	0.774	0.790	0.654	0.678	0.711	0.506
17	0.782	0.796	0.663	0.688	0.719	0.518
18	0.788	0.802	0.670	0.698	0.728	0.528
19	0.795	0.808	0.677	0.706	0.735	0.538
20	0.800	0.813	0.684	0.715	0.742	0.548
21	0.806	0.818	0.690	0.722	0.749	0.556
22	0.811	0.823	0.696	0.729	0.755	0.565
23	0.815	0.827	0.702	0.735	0.761	0.573
24	0.820	0.831	0.707	0.741	0.766	0.580
25	0.823	0.833	0.712	0.747	0.770	0.587
30		0.849	0.734		0.791	0.618
35		0.861	0.751		0.807	0.643
40		0.870	0.765		0.820	0.663
45		0.878	0.778		0.831	0.681
50		0.884	0.788		0.839	0.695

Using a normal approximation for the distribution of V_2/V_1 with the mean and the variance given by (4.1) and (4.2) respectively, the values of $\lambda_{.05}^{(1)}$ and $\lambda_{.01}^{(1)}$ are obtained and are given in columns 3 and 6 of Table 1. For comparison, the values of λ corresponding to $P\left(\frac{\delta^2}{2s^2} \leqslant \lambda\right) = 0.05$ and $P\left(\frac{\delta^2}{2s^2} \leqslant \lambda\right) = 0.01$, denoted by $\lambda_{.05}^{(0)}$ and $\lambda_{.01}^{(0)}$ respectively are also given in columns (4) and (7) of Table 1. These values are obtained from Hart's (1942) table of percentage points for δ^2/s^2. From Table 1 it is evident that Dixon's approximation is fairly accurate when N is greater than 15, particularly for 5% values. Since the exact evaluation becomes cumbersome as N increases we give only values of $\lambda_{.05}^{(1)}$ and $\lambda_{.01}^{(1)}$ obtained from Dixon's approximation for some selected values of $N > 25$.

Since the percentage points of V_3/V_2 may be quite useful in the variate difference method, the values of λ corresponding to $G_2(\lambda) = 0.05$ and $G_2(\lambda) = 0.01$ denoted by $\lambda_{.05}^{(2)}$ and $\lambda_{.01}^{(2)}$ respectively are obtained using the same procedure employed for V_2/V_1. The values of $\lambda_{.05}^{(2)}$ and $\lambda_{.01}^{(2)}$ are given in columns (2) and (4) of Table 2.

TABLE 2.   PERCENTAGE POINTS OF V_3/V_2

N	$\lambda_{.05}^{(2)}$		$\lambda_{.01}^{(2)}$	
	exact	approximate	exact	approximate
(1)	(2)	(3)	(4)	(5)
9	0.765		0.645	
10	0.779		0.679	
11	0.786		0.696	
12	0.800		0.708	
13	0.809		0.722	
14	0.816		0.734	
15	0.821		0.743	
16	0.828		0.752	
17	0.836		0.762	
18	0.841		0.770	
19	0.846		0.777	
20	0.851	0.863	0.783	0.813
21	0.855	0.867	0.789	0.817
22	0.859	0.870	0.794	0.822
23	0.863	0.873	0.800	0.826
24	0.867	0.876	0.806	0.830
25	0.870	0.879	0.812	0.834
30		0.890		0.849
35		0.899		0.861
40		0.906		0.871
45		0.912		0.879
50		0.917		0.885

The evaluation of the first two moments of V_3/V_2 through Dixon's method of approximation seems to be complicated. Therefore, we use the approximate formulas for the first two moments of the ratio of two random variables.

Then to $0(N^{-1})$ we find

$$E\left(\frac{V_3}{V_2}\right) = 1 + (\sigma_2^2 - \rho_2\sigma_2\sigma_3)/\sigma^4 \qquad \ldots \ (4.3)$$

and

$$\text{var}\left(\frac{V_3}{V_2}\right) = \sigma_3^2 + \sigma_2^2 - 2\rho_2\sigma_2\sigma_3/\sigma^4 \qquad \ldots \ (4.4)$$

since

$$E(V_2) = E(V_3) = \sigma^2. \qquad \ldots \ (4.5)$$

Now using (2.12) we obtain

$$E\left(\frac{V_3}{V_2}\right) = 1 - \frac{0.312}{N} \qquad \ldots \ (4.6)$$

and

$$\text{var}\left(\frac{V_3}{V_2}\right) = \frac{98}{900N}. \qquad \ldots \ (4.7)$$

Using now a normal approximation for the distribution of V_3/V_2 with the mean and the variance given by (4.6) and (4.7) respectively, the values of $\lambda_{.05}^{(2)}$ and $\lambda_{.01}^{(2)}$ are obtained and are given in columns (3) and (5) of Table 2. Since the approximate formulas are satisfactory only for large N, we tabulated the values for $N \geqslant 20$. The normal approximation seems to be quite satisfactory.

5. Example

To illustrate the use of the percentage points given in Tables 1 and 2 in the application of the variate difference method, we take the example of the yearly series of the quantity of meat consumed in the United States, 1919-1941 (Tintner, 1952, p. 320). The noncircular definition of the variance of the k-th difference is

$$V_k'' = \frac{\sum_{t=1}^{N-k} (\Delta^k x_t)^2}{(N-k)\binom{2k}{k}}. \qquad \ldots \ (5.1)$$

It is found that for the present data

$$V_0'' = 62.2517 \qquad V_1'' = 23.5864$$
$$V_2'' = 17.0411 \qquad V_3'' = 15.7392.$$

The problem now is to find the order of the finite difference k where the systematic part (or trend) of the time series is sufficiently eliminated. Now we have

$$\frac{\delta^2}{2s^2} = \frac{N}{N-1}\ \frac{V_1''}{V_0''} = 0.3789**$$

$$\frac{V_2''}{V_1''} = 0.7255**$$

and

$$\frac{V_3''}{V_2''} = 0.9236$$

where $N = 23$. Using the percentage points in Tables 1 and 2 which correspond to the circular statistic V_{k+1}/V_k, as approximations to the percentage points of the noncircular statistic V_{k+1}''/V_k'', we find that V_1''/V_0'' and V_2''/V_1'' are significant at 1% level but V_3''/V_2'' is not significant at 5% level. Therefore, the systematic part may be considered to have been sufficiently eliminated in the finite difference series of order $k = 2$.

References

Anderson, O. (1929): *Die Korrelationsrechnung in der Konjunkturforschung*, Frankfort.

Dixon, W. J. (1944): Further contributions to the problem of serial correlation. *Ann. Math. Stat.*, 15, 119-144.

Gurland, J. (1948): Inversion formulae for the distribution of ratios. *Ann. Math. Stat.*, 19, 228-237.

Hart, B. I. (1942): Significance levels for the ratio of the mean square successive difference to the variance. *Ann. Math. Stat.*, 13, 445-447.

Kamat, A. R. (1954): Distribution theory of two estimates for standard deviation based on second variate differences. *Biometrika*, 41, 1-11.

Koopmans, T. (1942): Serial correlation and quadratic forms in normal variables. *Ann. Math. Stat.*, 13, 14-33.

Tintner, G. (1940): *The Variate Difference Method*, Bloomington, Indiana.

———— (1952): *Econometrics*, John Wiley and Sons, Inc., New York.

———— (1955): The distribution of the variances of variate differences in the circular case. *Metron*, 17.

———— (1960): *Handbuch der Okonometrie*, Springer Verlag, Berlin.

Paper received : April, 1961.

LIMIT DISTRIBUTIONS OF THE CIRCLUAR SERIAL
CORRELATION COEFFICIENT

By ROY LEIPNIK

Michelson Laboratory, California

SUMMARY. Limit distributions of a renormalized circular serial correlation co-efficient from a circular version of a stationary Gaussian-Markov process are derived from expressions obtained by Madow (1945) and Leipnik (1947). The limit calculations are validated by summability arguments. Characteristic functions and moments are found. Iterated limit distributions, as correlation tends to 1, are also discussed. .

1. INTRODUCTION

The distribution of the serial correlation coefficient from a stationary Gaussian Markov process is extremely complicated (Leipnik, 1958b). The circular (periodic) modification, while somewhat simpler in appearance, remains rather opaque. A smoothed approximate distribution, whose first $\left(\dfrac{n}{2}\right)$ moments are correct, has been derived (Leipnik, 1947), but its overall accuracy is questionable.

The differences between distributions are often illuminated and sharpened by the study of limiting forms. A study of limiting forms of the circular distribution and the smoothed circular distribution seemed justified. The formal results were easily obtained, but justification of the steps was hard to find. The necessary tools were found in Hurwitz's theory of the consistency of Toeplitz methods of summability with Cesaro summability.

2. THE UHLENBECK-ORNSTEIN PROCESS AND ITS CIRCULAR MODIFICATION

For $0 < \rho < 1$, consider the Gaussian process $\{X(t),\ 0 \leqslant t < \infty\}$ such that $E[X(t_1)X(t_2)] = \rho^{|t_2 - t_1|}$, $E[X(t)] = 0$, known as the Uhlenbeck-Ornstein process. It is stationary and Markov; and conversely, the only such one-dimensional processes are of Uhlenbeck-Ornstein type. For each $h > 0$, the process $\{X_0^{(h)}, X_1^{(h)}, \ldots, X_n^{(h)}\}$, $X_j^{(h)} = X(jh)$ is a discrete stationary Gaussian-Markov process. The serial correlation coefficient

$$r_n^{(h)} = \frac{\sum\limits_{j=0}^{n-1} X_j^{(h)} X_{j+1}^{(h)}}{\sum\limits_{j=0}^{n-1} (X_j^{(h)})^2}$$

(sometimes defined to have $\frac{1}{2}[(X_0^{(h)})^2 + (X_n^{(h)})^2] + \sum\limits_{j=1}^{n-1}(X_j^{(h)})^2$ in the denominator) has been extensively studied as an estimator of ρ^h. Although the Von Neumann statistic $\sum\limits_{j=0}^{n-1}(X_{j+1}^{(h)} - X_j^{(h)})^2$ is now recognized to lead to better estimates of ρ, still $r_n^{(h)}$ holds much interest. Unfortunately, its distribution is excessively complicated (Leipnik, 1958b).

9

This difficulty is reduced by introducing, after Hotelling, a periodic (circular) process $\tilde{X}_j^{(h)}$ and associated coefficient. To arrive at these, note that $X_0^{(h)}$ and $X_{j+1}^{(h)} - \rho^h X_j^{(h)} = V_j^{(h)}$ are independent Gaussian with mean zero,

and
$$E[(V_j^{(h)})]^2 = \sigma^2(1 - \rho^{2h}) \text{ for } j = 0, ..., n-1.$$

Now define $\tilde{V}_1^{(h)}, ..., \tilde{V}_n^{(h)}$ to be independent Gaussian random variables with mean zero,

$$E[(\tilde{V}_j^{(h)})^2] = \tilde{\sigma}^2(h, \rho, n), \qquad j = 1, ..., n,$$

and define the $\tilde{X}_j^{(h)}$ by

$$\tilde{X}_{j+1}^{(h)} - \rho^h \tilde{X}_j^{(h)} = \tilde{V}_j^{(h)}, \qquad j = 0, ..., n-1 \text{ and } \tilde{X}_n^{(h)} = \tilde{X}_0^{(h)}.$$

It follows that
$$\tilde{X}_j^{(h)} = \frac{1}{1 - \rho^{nh}} \sum_{k=0}^{n-1} \tilde{V}_k^{(h)} \rho^{((j-k+n-1) \bmod n)h}.$$

Direct calculation yields

$$E[\tilde{X}_j^{(h)} \tilde{X}_k^{(h)}] = \frac{\tilde{\sigma}^2(h, \rho, n)}{(1 - \rho^{2h})(1 - \rho^{nh})} \; (\rho^{|j-k|h} + \rho^{(n-|j-k|)h}) \text{ for } 0 \leqslant j, \; k \leqslant n-1.$$

If in particular $\tilde{\sigma}^2(h, \rho, h)$ is taken to be $\sigma^2(1 - \rho^{2h})$,

then
$$E[\tilde{X}_j^{(h)} \tilde{X}_k^{(h)}] = \frac{\sigma^2}{1 - \rho^{nh}} \; (\rho^{|j-k|h} + \rho^{(n-|j-k|)h}),$$

a convenient form for most purposes. (For study of $\rho \rightarrow 1$, another choice of $\tilde{\sigma}^2(h, \rho, n)$ is preferable.)

The circular serial correlation coefficient

$$\tilde{r}_n^{(h)} = \frac{\sum\limits_{j=0}^{n-1} \tilde{X}_j^{(h)} \tilde{X}_{j+1}^{(h)} + \tilde{X}_{n-1}^{(h)} \tilde{X}_0^{(h)}}{\sum\limits_{j=0}^{n-1} (\tilde{X}_j^{(h)})^2}$$

is an estimator of ρ^h with a simpler distribution (Madow, 1945) than that of $r_n^{(h)}$.

The continuous version of $\tilde{X}_j^{(h)}$ is defined as follows : $\{\tilde{X}_T(t), \; 0 \leqslant t \leqslant T\}$ is a Gaussian stationary process with mean zero, and autocorrelation

$$E[\tilde{X}_T(t_1) \tilde{X}_T(t_2)] = \frac{\sigma^2}{1 - \rho^T} \; (\rho^{|t_2 - t_1|} + \rho^{T - |t_2 - t_1|}).$$

Thus
$$E[(\tilde{X}_T(T) - \tilde{X}_T(0))^2] = 0, \; \tilde{X}_T(T) = \tilde{X}_T(0)$$

with probability 1.

Note that as $T \rightarrow \infty$, the above autocorrelation tends for $\rho < 1$ to $\sigma^2 \rho^{|t_2 - t_1|}$, the autocorrelation of the Uhlenbeck-Ornstein, process. Clearly, for given n, h the

random variables $\tilde{X}_0^{(h)}, ..., \tilde{X}_{n-1}^{(h)}$ can be recovered from $\tilde{X}_T(t)$ by setting $T = nh$ and $\tilde{X}_j^{(h)} = \tilde{X}_T(jh)$, $j = 0, ..., n-1$. Henceforth this embedding of $\{\tilde{X}_j^{(h)}\}$ in $\{\tilde{X}_T(t)\}$ will be assumed. It will also be convenient to set $\beta = \log \rho$, so that

$$E[\tilde{X}_T(t_1)\tilde{X}_T(t_2)] = \sigma^2 \; \frac{\cosh\left(\frac{T}{2} - |t_2 - t_1|\right)\beta}{\sinh\frac{\beta T}{2}}.$$

3. DISTRIBUTION OF $\tilde{r}_n^{(h)}$

Madow (1945) showed how to derive the distribution of $\tilde{r}_n^{(h)}$ (see also Lehmann, 1947) which after simplification becomes

$$G_{n,h,\beta}(x) = P_r(\tilde{r}_n^{(h)} \leqslant x) = 1 - \frac{4\exp\left(\frac{-\beta h}{2}\right)\sinh\frac{\beta n h}{2}}{n(1-\exp(-\beta h))} \qquad \dots \quad (3.1)$$

$$\sum_{k=1}^{\rho_n^{(x)}} (-1)^{k-1}\cos\frac{\pi k}{n}\left(\frac{1-\cos\frac{2\pi k}{n}}{\cosh(\beta h)-\cos\frac{2\pi k}{n}}\right)\left(\frac{\cos\frac{2\pi k}{n}-x}{\cosh(\beta h)-x}\right)^{(n-3)/2},$$

where $\rho_n^{(x)}$ is the first non-negative integer j such that

$$\cos\frac{2\pi j}{n} \geqslant x > \cos\frac{2\pi(j+1)}{n}.$$

An approximation $\bar{G}_{n,h,\beta}$ to $G_{n,h,\beta}$ with the same first $\binom{n}{2}$ moments is

$$\bar{G}_{n,h,\beta}(x) = \frac{\Gamma\left(\frac{n}{2}+1\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n}{2}+\frac{1}{2}\right)} \cdot \quad \text{(Leipnik, 1947)} \qquad \dots \quad (3.2)$$

$$\int_{-1}^{x} (1-t^2)^{(n-1)/2}(1+\exp(-2\beta h)-2t\exp(-\beta h))^{-n/2}dt.$$

Henceforth (3.1) and its related forms will be known as exact, whereas (3.2) and its related forms will be known as smoothed.

4. A SEQUENCE OF ESTIMATORS OF ρ^T

There is a way of renormalizing a sequence $\{u_n\}$ of statistics which may force it to converge to a constant. Suppose the distributions of $\{u_n\}$ contain a parameter γ, and that u_n is, in some sense, an estimator of $\alpha(\gamma, n)$. If $\beta(\alpha(\gamma, n), n) = \gamma$, then we say that $\beta(u_n, n)$ is an estimate-normalized sequence of statistics for γ. Of course, this terminology is only suggestive. Since $\tilde{r}_n^{(h)}$ is an estimator of $\rho^h = \exp(-\beta h)$,

the sequence $(\tilde{r}_n^{(h)})$ is an estimate-renormalized sequence for $\exp(-\beta nh)$. If now $\tilde{X}_j^{(h)}$ is imbedded in $\tilde{X}_T(t)$ with $T = nh$ in the manner shown in Section 2, then $\{s_n\}$ is an estimate-renormalized sequence for $\exp(-\beta T) = \rho^T$,

where $$s_n = (\tilde{r}_n^{(T/n)})^n. \qquad \ldots \text{(4.1)}$$

Since $\tilde{r}_n^{(h)}$ is not a positive random variable, s_n is positive only for n even.

If $$K_{n,T,\beta}(y) = Pr[s_n \leqslant y], \qquad \ldots \text{(4.2)}$$

then for n even $K_{n,T,\beta}(y) = \begin{cases} 0, \ y < 0 \\ G_{n,T/n,\beta}(y^{1/n}) - G_{n,T/n,\beta}(-y^{1/n}), \ 0 \leqslant y \leqslant 1 \end{cases}$

and for n odd $\qquad K_{n,T,\beta}(y) = \begin{cases} G_{n,T/n,\beta}(-y)^{1/n}, \ y < 0 \\ G_{n,T/n,\beta}(y^{1/n}), \ 0 \leqslant y \leqslant 1 \end{cases}$

In particular $\qquad K_{n,T,\beta}(0) = \begin{cases} 0 & \text{for } n \text{ even} \\ G_{n,T/n,\beta(0)} & \text{for } n \text{ odd} \end{cases}$

If $\lim\limits_{n} K_{n,T,\beta}(0) = 0$ and $\lim\limits_{n} G_{n,T,\beta}(y^{1/n}) = K_{T,\beta}(y)$ exists and is a distribution for $0 \leqslant y \leqslant 1$, then $K_{T,\beta}$ will be the limiting distribution of $\{s_n\}$.

Similarly, a smoothed distribution $\bar{K}_{n,T,\beta}$ can be defined in terms of $\bar{G}_{n,T/n,\beta}$. If $\lim\limits_{n} \bar{K}_{n,T,\beta}(0) = 0$ and $\lim\limits_{n} \bar{G}_{n,T/n,\beta}(y^{1/n}) = \bar{K}_{T,\beta}(y)$ exists for $0 < y \leqslant 1$, and $\bar{K}_{T,\beta}$ is a distribution, it is the limiting smoothed distribution.

$$5. \quad \lim_{n} G_{n,T/n,\beta}(0) = 0$$

Let $$G_{n,\delta} = G_{n,T/n,\beta}(0) = \frac{1 - 4\exp\left(\dfrac{\delta}{-n}\right)\sinh\delta}{n\left(1 - \exp\left(\dfrac{2\delta}{-n}\right)\right)}. \qquad \ldots \text{(5.1)}$$

$$\sum_{k=1}^{n/4}(-1)^{k-1}\cos\frac{\pi k}{n}\left(\frac{1 - \cos\dfrac{2\pi k}{n}}{\cosh\dfrac{2\delta}{n} - \cos\dfrac{2\pi k}{n}}\right)\left(\frac{\cos\dfrac{2\pi k}{n}}{\cosh\dfrac{2\delta}{n}}\right)^{(n-3)/2},$$

where $\delta = \dfrac{\beta T}{2}$.

Formally taking $n \to \infty$, writing $\sum\limits_{k=1}^{\infty}(-1)^{k-1} = \frac{1}{2}$, and using the identity

$$z^2\sum_{k=1}^{\infty}\frac{(-1)^{k-1}}{z^2+\pi^2k^2} = \frac{1}{2}\left(1 - \frac{z}{\sinh z}\right),$$

[Whittaker and Watson (1948, p.136)] we quickly find $\lim\limits_{n} G_{n,\delta} = 0$. The justification obviously calls for a summability argument.

If $T = [t_{k,n}]$ is an infinite matrix, the sequence $\left[\sum_{k=1}^{\infty} t_{k,n} a_k\right]$ is called the Toeplitz sum sequence of the series $\sum_{k=1}^{\infty} a_k$, and its limit, if it exists, is called the Toeplitz sum $(T) \sum_{k=1}^{\infty} a_k$. In particular, if $t_{,n} = 1-k/n$, $1 \leqslant k \leqslant n$; $t_{,n} = 0$, $k > n$, the Cesaro sum is obtained, denoted by

$$(C) \sum_{K=1}^{\infty} a_K.$$

The result $\qquad (C) \sum_{K=1}^{\infty} (-1)^{K-1} = \tfrac{1}{2}$

is classical. The (T) sum is said to be consistent with the (C) sum if

$$(T) \sum_{K=1}^{\infty} a_K = (C) \sum_{K=1}^{\infty} a_K \text{ whenever } (C) \sum_{K=1}^{\infty} a_K \text{ is finite.}$$

The Hurwitz conditions (see Moore, 1938) for (T) sum to be consistent with (C) sum are

$$(H_1) \lim_{K \to \infty} t_{K,n} = 0; \quad (H_2) \lim_{K \to \infty} t_{K,n} = 1; \quad (H_3) \sup_n \sum_{K=1}^{\infty} K |\Delta_K^2 t_{K,n}| < \infty,$$

where $\qquad \Delta_K^2 t_{K,n} = t_{K+2,n} - 2t_{K+1,n} + t_{K,n}.$

Let $t_{K,n} = \left(\cos \frac{2\pi K}{n}\right)^{b_n}$, $b_n = \frac{n-3}{2}$, $1 \leqslant K \leqslant \left(\frac{n}{4}\right)$; $t_{K,n} = 0$, $K > \left(\frac{n}{4}\right)$.

(H_1) is obviously satisfied, and (H_2) is easily checked since

$$\lim_n t_{K,n} = \left(\lim_{n \to \infty} \left(\cos \frac{2\pi K}{n}\right)^n\right)^{\frac{1}{2}},$$

and $\qquad 1 - \frac{2\pi^2 K^2}{n} \leqslant \left(1 - \frac{2\pi^2 K^2}{n^2}\right)^n$

$$\leqslant \left(\cos \frac{2\pi K}{n}\right)^n \leqslant 1.$$

As for (H_3), note that

$$\frac{d^2}{dx^2}(\cos ax)^n = a^2 b_n (\cos ax)^{b_n-2} [(b_n-1) \sin^2 ax - \cos^2 ax]$$

is positive for $x > a^{-1} \sin^{-1} b_n^{-\frac{1}{2}}$ and negative for $x < a^{-1} \sin^{-1} b_n^{-\frac{1}{2}}$. Thus $(\cos ax)^{b_n}$ is convex in $(0, a^{-1} \sin^{-1} b_n^{-\frac{1}{2}})$ and concave in $\left(a^{-1} \sin^{-1} b_n^{-\frac{1}{2}}, a^{-1} \frac{\pi}{2}\right)$ for $b_n \geqslant 1$. It follows that there is an integer d_n,

$$\left|d_n - \frac{n}{2\pi} \sin^{-1} (b_n)^{-\frac{1}{2}}\right| < 1,$$

such that
$$\Delta^2 \left(\cos \frac{2\pi K}{n} \right)^{b_n} \leqslant 0 \text{ for } 1 \leqslant K < d_n$$

and
$$\Delta^2 \left(\cos \frac{2\pi K}{n} \right)^{b_n} \geqslant 0 \text{ for } d_n < K \leqslant \left(\frac{n}{4} \right).$$

Application of the partial summation formula
$$\sum_{K_1}^{K_2} K \Delta^2 t_{K, n} = (K_2+1)\Delta t_{K_2+1,n} - K_1 \Delta t_{K_1,n} - t_{K_2+2,n} + t_{K_1+1,n},$$

easily proved by induction, yields

$$\sum_{K=1}^{(n/4)} K \, |\Delta^2 t_{K, n}| = - \sum_{K=1}^{d_n-1} K \, \Delta^2 t_{K, n} + d_n |\Delta^2 t_{d_n,n}| + \sum_{K=d_n+1}^{(n/4)} K \, \Delta^2 t_{K, n}$$

$$= d_n [|\Delta^2 t_{d_n,n}| - \Delta t_{d_n,n} - \Delta t_{d_n+1,n}] + [2 t_{d_n+1,n} - t_{1,n}]$$

$$\leqslant 2 d_n \max [|\Delta t_{d_n,n}|, \; |\Delta t_{d_n+1,n}|] + 3$$

since $|t_{K, n}| \leqslant 1$ for $n > \dfrac{-t}{s}$. Clearly, both of these first differences are bounded above

by the value $|\dfrac{d}{dx} (\cos ax)^{b_n}|$ at $x = a^{-1} \sin^{-1} b_n^{-\frac{1}{2}}$, which equals $a b_n^{\frac{1}{2}} (\cos \sin^{-1} b_n^{-\frac{1}{2}})^{b_n-1}$.

Hence
$$d_n \max [|\Delta C_{d_n,n}|, \; |\Delta C_{d_n,1 n}|] \leqslant \frac{\sin^{-1} b_n^{-\frac{1}{2}}}{b_n^{-\frac{1}{2}}} (1-b_n^{-1})^{(b_n-1)/2}$$

which tends to 1 as $n \to \infty$, and therefore is bounded proving (H_3).

Now consider the Toeplitz matrix defined by $t'_{K, n} = \left(\cos \dfrac{\pi K}{n} \right) t_{K, n}$. If $\{a_K\}$ is bounded, then

$$\overline{\lim_n} \, | \sum_{K=1}^{\infty} (t_{,n} - t'_{K, n}) a_K | \leqslant \sup_K |a_K|. \; \overline{\lim_n} \sum_{K=1}^{(n/4)} \left(1-\cos \frac{\pi K}{n} \right) t_{D, n}.$$

Note that $1-\cos 2\pi x [\exp (2\pi^2 x^2)]$ is 0 at $x = 0$ and its derivative

$$2\pi \cos 2\pi x \, \exp(2\pi^2 x^2)(\tan 2\pi x - 2\pi x) > 0 \text{ for } x < \frac{1}{4}.$$

Hence
$$\cos 2\pi x \leqslant \exp (-2\pi^2 x^2), \text{ for } x < \frac{1}{4}.$$

For
$$n^{7/12} < K \leqslant \left(\frac{n}{4} \right), \; \cos \frac{2\pi K}{n} < \cos 2\pi n^{-5/12} \leqslant \exp (-2\pi^2 n^{-5/6}),$$

$$\left(\cos \frac{2\pi K}{n} \right)^{n/2} \leqslant \exp (-\pi^2 n^{1/6}),$$

so that
$$\sum_{K > n^{7/12}} t_{K,n} \cdot \left(1-\cos \frac{\pi K}{n} \right) \leqslant \frac{n}{4} \exp (-\pi^2 n^{1/6})(\cos 2\pi n^{-5/12})^{-3/2} \to 0.$$

On the other hand,

$$\left(\cos \frac{2\pi K}{n} \right)^{n/2} \leqslant 1,$$

so
$$\sum_{K \,\leqslant\, n^{7/12}} t_{K,\,n} \left(1 - \cos \frac{\pi K}{n} \right) \leqslant \sum_{K \,\leqslant\, n^{7/12}} \left(\cos \frac{2\pi K}{n} \right)^{-3/2} \left(1 - \cos \frac{\pi K}{n} \right)$$

$$\leqslant \sum_{K \,\leqslant\, n^{7/12}} \left(1 - \frac{2\pi^2 K^2}{n^2} \right)^{-3/2} \frac{\pi^2 K^2}{2n^2}$$

$$\leqslant 10 \sum_{K \,\leqslant\, n^{7/12}} \left(1 + \frac{3\pi^2 K^2}{n^2} \right) \frac{\pi^2 K^2}{2n^2}$$

$$\leqslant 5\pi^2 n^{-1/4} + 15\pi^4 n^{-13/12} \quad \text{for} \quad n > 5.$$

Hence
$$\varlimsup_{n} \left| \sum_{K=1}^{\infty} (t_{K,n} - t'_{K,n}) a_K \right| = 0,$$

and the modified summability method is also consistent with Cesaro summability.

Thus
$$\lim_{n} \sum_{K=1}^{(n/4)} (-1)^{K-1} t'_{kn} = (C) \sum_{K=1}^{\infty} (-1)^{K-1} = \tfrac{1}{2}.$$

Now

$$\sum_{K=1}^{(n/4)} (-1)^{K-1} t'_{K,n} \left(\frac{1 - \cos \frac{2\pi K}{n}}{\cosh \frac{2\delta}{n} - \cos \frac{2\pi K}{n}} \right) = \sum_{K=1}^{(n/4)} (-1)^{K-1} t'_{K,n} - \left(\cosh \frac{2\delta}{n} - 1 \right) n^2$$

$$\times \sum_{K=1}^{(n/4)} (-1)^{K-1} t'_{K,n} \left[n^2 \left(\cosh \frac{2\delta}{n} - \cos \frac{2\pi K}{n} \right) \right]^{-1}. \qquad \ldots \ (5.2)$$

At this point it is helpful to invoke a lemma :

Lemma : *If $\rho_n \to \infty$, $f_{K,n}$ is a non-negative double sequence such that for some n_0, $f_{K,n} \leqslant f_{K+1,n}$ for $n > n_0$, $1 \leqslant K < n$, $\lim_{n \to \infty} f_{Kn} = f_K$ exists for all K, and $\sum_{K=1}^{\infty} (-1)^{K-1} f_K$ converges, then*

$$\lim_{n \to \infty} \sum_{K=1}^{\rho_n} (-1)^{K-1} f_{K,n} = \sum_{K=1}^{\infty} (-1)^{K-1} f_K.$$

For the proof, see Leipnik (1958a). Now take

$$f_{K,n} = t'_{K,n} \left[n^2 \left(\cosh \frac{2\delta}{n} - \cos \frac{2\pi K}{n} \right) \right]^{-1},$$

and note that for $n \geqslant 3$, $f_{K,n}$ is a product of monotone non-increasing expressions in K, that $\rho_n = (n/4) \to \infty$, that $\lim_n f_{K,n} = \lim_n n^{-2} \left(\cosh \frac{2\delta}{n} - \cos \frac{2\pi K}{n} \right)^{-1} = (2\delta^2 + 2\pi^2 K^2)^{-1}$,

and that $\sum_{K=1}^{\infty} (-1)^{K-1} (2\delta^2 + 2\pi^2 K^2)^{-1}$ converges. Hence (5.2) converges to

$$\tfrac{1}{2} - \delta^2 \sum_{K=1}^{\infty} \frac{(-1)^{K-1}}{\delta^2 + \pi^2 K^2} = \frac{\delta}{2 \sinh \delta}.$$

Comparison with (5.1) shows that $G_{n,\delta}$ converges to

$$1-4\sinh\delta\,\frac{\delta}{2\sinh\delta}\,\lim_n\,\frac{\exp\left(\dfrac{\delta}{-n}\right)}{n\left(1-\exp\left(\dfrac{-2\delta}{n}\right)\left(\cosh\dfrac{2\delta}{n}\right)^{(n-3)/2}\right)}=0,$$

as stated. Since $G_{n'T/n,\beta}$ is a cumulative distribution, it follows that $\lim\limits_{n\to\infty} G_{n,T/n,\beta}$ $(-\xi_n)=0$ for all non-negative sequences $\{\xi_n\}$. From (4.2) $\lim\limits_{n} K_{n,T,\beta}\ (y)=0$ both for n even and n odd, and $\lim\limits_{n} K_{n,T\beta}\ (y)=\lim\limits_{n} G_{n\,T/n,\beta}\ (y^{1/n})$ for $0\leqslant y<1$, if the latter exists.

<div align="center">6.   Limiting exact distribution</div>

The analysis of

$$G_{n,T/n,\beta}(y^{1/n})=\frac{1-4\exp\left(-\dfrac{\delta}{n}\right)\sinh\delta}{n\left(1-\exp\left(-\dfrac{2\delta}{n}\right)\right)}$$

$$\times\sum_{K=1}^{\rho_n(y^{1/n})}(-1)^{K-1}\cos\frac{\pi K}{n}\left(\frac{\cos\dfrac{2\pi K}{n}-y^{1/n}}{\cosh\dfrac{2\delta}{n}-y^{1/n}}\right)^{(n-3)/2}$$

$$\times\left(1-\frac{\cosh\dfrac{2\delta}{n}-1}{\cosh\dfrac{2\delta}{n}-\cos\dfrac{2\pi K}{n}}\right)\qquad\qquad\dots\ (6.1)$$

is simpler than that of $G_{n,\delta}$. Recall that

$$1\geqslant\cos\left(\frac{2\pi}{n}\,\rho_n(y^{1/n})\right)\geqslant y^{1/n}>\cos\frac{2\pi}{n}\,(\rho_n(y^{1/n})+1)$$

so that

$$\frac{2\pi}{n}\,\rho_n(y^{1/n})\leqslant\cos^{-1}y^{1/n}<\frac{2\pi}{n}\,(\rho_n(y^{1/n})+1).$$

Now $n\cos^{-1}y^{1/n}\to\infty$ for $0<y<1$, since $\lim\limits_{x\to0}\dfrac{\cos^{-1}y^x}{x}=\lim\limits_{x\to0+}\dfrac{y^x\log 1/y}{1-y^{2x}}=\infty$

Also, $$\cos\frac{\pi K}{n}\left(\frac{\cos\dfrac{2\pi K}{n}-y^{1/n}}{\cosh\dfrac{2\delta}{n}-y^{1/n}}\right)^{(n-3)/2}$$

and $$\cos\frac{\pi K}{n}\left(\frac{\cos\dfrac{2\pi K}{n}-y^{1/n}}{\cosh\dfrac{2\delta}{n}-y^{1/n}}\right)\cdot\frac{\left(\cosh\dfrac{2\delta}{n}-1\right)}{\left(\cosh\dfrac{2\delta}{n}-\cos\dfrac{2\pi K}{n}\right)}$$

are decreasing in K and tend respectively to

$$\exp\left(\frac{\delta^2+\pi^2K^2}{\log y}\right)\text{ and }\exp\left(\frac{\delta^2+\pi^2K^2}{\log y}\right)\cdot\frac{\delta^2}{\delta^2+\pi^2K^2}.$$

Application of the Lemma to both parts of the above expression yields

$$K_\delta(y) = \lim_n G_{n,T/n,\beta}(y^{1/n}) = 1 - \frac{2\sinh\delta}{\delta} \sum_{K=1}^{\infty} (-1)^{K-1} \frac{\pi^2 K^2}{\delta^2 + \pi^2 K^2} \exp\left(\frac{\delta^2 + \pi^2 K^2}{\log y}\right) .$$

$$\text{for} \quad 0 < y < 1. \qquad\qquad ... \quad (6.2)$$

Trivially, $K_\delta(1) = \lim_n G_{,T/n,\beta}(1) = 1$, so that K_δ is indeed a distribution, continuous except perhaps at $0, 1$. Now

$$M_\delta(z) = K_\delta((\exp z^{-1})^{-1}) = \begin{cases} 1 - \dfrac{2\sinh\delta}{\delta} \sum\limits_{K=1}^{\infty} (-1)^{K-1} \dfrac{\pi^2 K^2}{\delta^2 + \pi^2 K^2} e^{-(\pi^2 K^2 + \delta^2)z}, & z < 0 \\[2ex] 0, & z \leqslant 0. \end{cases}$$

Since

$$\sum_{K=1}^{\infty} \sup_{z > C} \frac{\pi^2 K^2}{\delta^2 + \pi^2 K^2} e^{-(\pi^2 K^2 + \delta^2)z} \leqslant \sum_{K=1}^{\infty} e^{-(\pi^2 K^2 + \delta^2)C} \leqslant e^{-\delta^2}$$

$$\text{for each } C > 1,$$

it follows that summation over K and limits as $z \to \infty$ can be interchanged. Hence $\lim\limits_{y \to 1^-} K_\delta(y) = \lim\limits_{z \to \infty} M_\delta(z) = 1$. The continuity at other end is more difficult to obtain.

Consider the summability method defined by $\lim\limits_{z \to 0^+} \sum\limits_{K=1}^{\infty} a_K e^{-\lambda_K z}$. Garabedian (1931) showed that if

$$\lim_{K \to \infty} \frac{\lambda_K}{\log K} = \infty \quad \text{and for some } a > 0, \ \lim_{K \to \infty} \frac{\lambda_K}{K^a} = 0,$$

then

$$\lim_{z \to 0^+} \sum_{K=1}^{\infty} a_K e^{-\lambda_K z} = (C) \sum_{K=1}^{\infty} a_K.$$

Now for $\lambda_K = \pi^2 K^2$, these conditions are certainly satisfied.

Hence

$$\lim_{z \to 0^+} \sum_{K=1}^{\infty} (-1)^{K-1} \frac{\pi^2 K^2}{\delta^2 + \pi^2 K^2} e^{-\pi^2 K^2 z}$$

$$= (C) \sum_{K=1}^{\infty} (-1)^{K-1} \frac{\pi^2 K^2}{\delta^2 + \pi^2 K^2} = \frac{\delta}{2\sinh\delta} ,$$

and thus

$$\lim_{z \to 0^+} M_\delta(z) = \lim_{y \to 0^+} K_\delta(y) = 0.$$

The c.f. of M_δ is not hard to obtain. Let

$$\chi_\delta(t) = \int_0^{\infty} e^{itz} \, dM_\delta(z) = \lim_{\varepsilon \to 0^+} \int_\varepsilon^{\infty} e^{itz} \, dM_\delta(z) = 2 \frac{\sinh\delta}{\delta} \lim_{\varepsilon \to 0^+} \sum_{K=1}^{\infty} (-1)^{K-1}$$

$$\times \frac{\pi^2 K^2}{\pi^2 K^2 + \delta^2 - it} e^{-\varepsilon(\pi^2 K^2 + \delta^2 - it)} .$$

By an argument like that of the last paragraph,

$$\chi_\delta(t) = \frac{\sinh \delta}{\delta} \cdot \frac{\sqrt{\delta^2 - it}}{\sinh \sqrt{\delta^2 - it}}.$$

It is easy to show that

$$\chi_\delta(tT) = \sqrt{1 - \frac{2it}{\beta\delta}} \, (1 - e^{-2\delta}) \left(1 - \exp\left(-2\delta \sqrt{1 - \frac{2it}{\beta\delta}} \right) \right)^{-1}$$

$$\times \exp\left(\delta \left(1 - \sqrt{1 - \frac{2it}{\beta\delta}} \right) \right).$$

Since

$$\delta \left(1 - \sqrt{1 - \frac{2it}{\beta\delta}} \right) = \frac{2it/\beta}{1 + \sqrt{1 - \frac{2it}{\beta\delta}}},$$

clearly,

$$\lim_{\delta \to \infty} \chi_\delta(tT) = e^{it/\beta} . \qquad \ldots \ (6.3)$$

Thus if u_δ is a random variable with distribution K_δ and $v_\delta = (\exp v_\delta^{-1})^{-1}$, then v_δ is a random variable with distribution M_δ, and $W_\delta = Tv_\delta$ has a c.f. which tends to $e^{it/\beta}$, the c.f. of $\frac{1}{\beta}$, as $T \to \infty$. Hence in a weak sense $(r_n^{(T/n)})^{n/T}$ is an estimator of $\rho = e^{-\beta}$ for n and T large.

The distributions K_δ and M_δ are reminiscent of the Kolmogorov-Smirnov distribution, though slightly more complicated. The mean and variance of M_δ are easily calculated from χ_δ as

$$\frac{1}{2\delta} \left(\frac{1}{\tanh \delta} - \frac{1}{\delta} \right) \text{ and } \frac{1}{4\delta^2} \left(\frac{1}{\tanh^2 \delta} + \frac{1}{\delta \tanh \delta} - \frac{2}{\delta^2} - 1 \right).$$

7. $\lim_{n} \ \bar{G}_{n,T/n,\beta} \ (0) = 0$

A smoothing approximation $\bar{G}_{n,h,\beta}$ to $G_{n,h,\beta}$ was derived by Leipnik (1947) and Quenouille (1949). The expression is

$$\bar{G}_{n,h,\beta} (x) = \frac{\Gamma\left(\frac{n}{2} + 1 \right)}{\Gamma\left(\frac{1}{2} \right) \Gamma\left(\frac{n}{2} + \frac{1}{2} \right)} \cdot \int_{-1}^{x} (1 - t^2)^{(n-1)/2} (1 + \rho^{2h} - 2\rho^h t)^{-n/2} dt. \qquad \ldots \ (7.1)$$

To find the limiting renormalized distribution, the procedure of Sections 5 and 6 is imitated. Note that

$$\overline{G}_{n,\delta} = \overline{G}_{n,T/n,\beta}(0) = \frac{\int_0^1 (1-t^2)^{(n-1)/2} \left(1+\exp\left(\frac{-4\delta}{n}\right)+2\exp\left(\frac{-2\delta}{n}\right)t\right)^{-n/2} dt}{2\int_0^1 (1-t^2)^{(n-1)/2} dt}$$

$$\leqslant \frac{1}{2} \sup_{0\leqslant t\leqslant 1} \left[\left(1+\exp\left(\frac{-4\delta}{n}\right)+2\exp\left(\frac{-2\delta}{n}\right)t\right)^{-n/2}\right]$$

$$\leqslant \frac{1}{2}\left(1+\exp\left(\frac{-4\delta}{n}\right)\right)^{-n/2} \leqslant \frac{1}{2}\left(2-\frac{4\delta}{n}\right)^{-n/2},$$

so that

$$\lim_n \overline{G}_{n,\delta} = 0.$$

8. THE LIMITING SMOOTHED DISTRIBUTION

As before, let $\overline{K}_{n,\delta}(y) = \overline{G}_{n,T/n,\beta}(y^{1/n}) = \overline{G}_{n,\delta} + \int_0^y \phi_n(t)dt$,

where

$$\phi_n(t) = \frac{\Gamma\left(\frac{n}{2}+1\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n}{2}+\frac{1}{2}\right)}(1-t^{n/2})^{(n-1)/2}$$

$$\times \left(1+\exp\left(-\frac{4\delta}{n}\right)-2\exp\left(-\frac{2\delta}{n}\right)t^{1/n}\right)^{-n/2}\frac{1}{n}t^{(1/n)-1}. \qquad \ldots \quad (8.1)$$

A limit theorem for Lebesgue integration (Scheffe, 1947) states that if ϕ is non-negative measurable on $[0, 1]$, $\lim_n \phi_n = \phi$ exists,

$$\lim_n \int_0^1 \phi_n(t)dt = \int_0^1 \phi(t)dt = 1,$$

then

$$\lim_n \int_A \phi_n(t)dt = \int_A \phi(t)dt$$

for each measurable set A.

Now

$$\overline{G}_{n,T/n,\beta}(1) = 1 = \overline{G}_{n,\delta} + \int_0^1 \phi_n(t)dt,$$

so

$$1 = \lim_n \overline{G}_{n,\delta} + \lim_n \int_0^1 \phi_n(t)dt = \lim_n \int_0^1 \phi_n(t)dt.$$

Also,

$$\lim_n \frac{\Gamma\left(\frac{n}{2}+\frac{1}{2}\right)}{\left(\frac{n}{2}\right)^{\frac{1}{2}}\Gamma\left(\frac{n}{2}+\frac{1}{2}\right)} = 1, \quad \lim_n t^{1/n-1} = t^{-1}$$

for $\quad t > 0, \lim\limits_{n} (2n)^{-\frac{1}{2}} (1-t^{2/n})^{-\frac{1}{2}} = (-4 \log t)^{-\frac{1}{2}} \quad$ for $\quad 0 < t < 1,$

$$\lim\limits_{n} \left(\frac{1-t^{2/n}}{1+\exp\left(\frac{4\delta}{-n}\right)-2\exp\left(\frac{2\delta}{-n}\right)t^{1/n}} \right)^{n/2}$$

$$= \lim\limits_{n} \left[\exp\left(\frac{2\delta}{n}\right) \left(\frac{1-t^{2/n}}{2\left(\cosh\frac{2\delta}{n}-t^{1/n}\right)} \right) \right]^{n/2} = e\delta \, t^{\frac{1}{4}}e\delta^2/\log t,$$

so that $\quad \phi(t) = \lim\limits_{n} \phi_n(t) = \dfrac{e^\delta}{\sqrt{4\pi}} \, t^{-\frac{1}{4}}(-\log t)^{-\frac{1}{2}} \, e\delta^2/\log t \,.$ $\qquad$ ... (8.2)

Note that

$$\int_0^1 t^p\phi(t)dt = -\frac{2e^\delta}{\sqrt{4\pi}} \int_0^\infty \tau^{-2} \exp\left(-\delta^2\tau^2-\frac{p+\frac{1}{4}}{\tau^2} \right) d\tau = \frac{\exp\left(\delta(1-\sqrt{4p+1})\right)}{\sqrt{4p+1}} \,,$$

so $\qquad\qquad\qquad\qquad\qquad \int_0^1 \phi(t)dt = 1,$

and the p-th moment of $\quad \bar{K}_\delta(y) = \dfrac{e^\delta}{\sqrt{4\pi}} \int_0^y t^{-\frac{3}{4}} (-\log t)^{-\frac{1}{2}} \, e\delta^2/\log t \; dt \qquad$ ... (8.3)

is $\qquad\qquad\qquad\qquad \dfrac{\exp\left(\delta(1-\sqrt{4p+1})\right)}{\sqrt{4p+1}}.$

The transformed distribution

$$\bar{M}_\delta(y) = \bar{K}_\delta((\exp y^{-1})^{-1}) = \frac{e^\delta}{\sqrt{4\pi}} \int_0^y s^{-3/2} \exp\left(-\frac{1}{4s}-\delta^2 s \right) ds \qquad \text{... (8.4)}$$

$$\text{for } 0 \leqslant y < \infty.$$

The c.f. χ_δ of $\bar{M}_\delta$ is easily found as

$$\bar{X}_\delta(t) = \frac{e^\delta}{\sqrt{\pi}} \int_0^\infty \tau^{-2} \exp\left(-\frac{1}{4\tau^2}-(\delta^2-it)\tau^2 \right) d\tau = \exp\left(\delta\left(1-\sqrt{1-\frac{it}{\delta^2}} \right) \right).$$

$$\text{... (8.5)}$$

The mean and variance of $\bar{M}_\delta$ are $\dfrac{1}{2\delta}$ and $\dfrac{1}{4\delta^3}$ respectively.

Clearly, $\bar{X}_\delta(tT) = \exp\left(\delta\left(1-\sqrt{1-\frac{2it}{\beta\delta}} \right) \right) \to \exp\left(\frac{it}{\beta} \right)$ as $T\to\infty.$

Thus although χ_δ and $\bar{x}_\delta$ have a quite different appearance, the limits as $T\to\infty$ of $\bar{\chi}_\delta(tT)$ and $\chi_\delta(tT)$ are the same, the c.f. of the constant $\dfrac{1}{\beta}.$

9. LIMITING DISTRIBUTIONS AS $\delta \to 0$

Formally, the above distributions and c.f.'s have limits at $\delta \to 0$. For the exact distributions,

$$
\left.
\begin{aligned}
K_0(y) &= \lim_{\delta \to 0} K_\delta(y) = 1 - 2 \sum_{K=1}^{\infty} (-1)^{K-1} \exp\left(\frac{\pi^2 K^2}{\log y}\right) \\
M_0(y) &= 1 - 2 \sum_{K=1}^{\infty} (-1)^{K-1} e^{-\pi^2 K^2 y} \\
\chi_0(t) &= \frac{\sqrt{-it}}{\sinh\sqrt{-it}}
\end{aligned}
\right\} \qquad \dots \quad (9.1)
$$

For the smoothed distributions,

$$
\left.
\begin{aligned}
\bar{K}_0(y) &= \frac{1}{\sqrt{4\pi}} \int_0^y t^{-\frac{1}{2}} (-\log t)^{-\frac{1}{2}} dt \\
\bar{M}_0(y) &= \frac{1}{\sqrt{4\pi}} \int_0^y s^{-3/2} \exp\left(\frac{-1}{4s}\right) ds \\
\bar{\chi}_0(t) &= \exp\left(\sqrt{-it}\right)
\end{aligned}
\right\} \qquad \dots \quad (9.2)
$$

Curiously enough, M_0 has moments of all orders, while the p-th moment of $\bar{M}_0$ is $\dfrac{\Gamma(\frac{1}{2}-p)}{\Gamma(\frac{1}{2})}$ for $p < \frac{1}{2}$, infinite for $p \geqslant \frac{1}{2}$. It is amusing to note that

$$
\bar{N}_0(y) = \bar{M}_0(2y^2)^{-1}) = \sqrt{\frac{2}{\pi}} \int_y^{\infty} e^{-u^2/2} du
$$

is the positive normal distribution.

There is a reason to expect such limiting distributions to exist as $\delta \to 0$. In Section 2, a set of independent random variables of variance $\tilde{\sigma}^2(h, \rho, n)$ was introduced, and a related set $X_j^{(h)}$ of random variables of autocorrelation

$$
\frac{\tilde{\sigma}^2(h, \rho, n)}{(1-\rho^{2h})(1-\rho^{nh})} \left(\rho^{|j-k|h} + \rho^{(n-|-k|)h}\right)
$$

was defined. For simplicity, the agreement $\tilde{\sigma}^2(h, n, \rho) = \sigma^2(1-\rho^{2h})$ was made, and the distribution of $\tilde{r}_n^{(h)}$ was written down. Since $\tilde{r}_n^{(h)}$ is homogeneous of degree zero, its distribution is independent of the choice of $\tilde{\sigma}^2(h, \rho, n)$. If the choice $\tilde{\sigma}^2(h, n, \rho) = \sigma^2\dfrac{(1-\rho^{2h})}{1-\rho^2}$ is taken, so that the autocorrelation of the corresponding continuous process $\{\tilde{X}_T(t)\}$ becomes

$$
E[\tilde{X}_T(t_1)\tilde{X}_T(t_2)] = \frac{\sigma^2(\rho^{|t_2-t_1|} + \rho^{T-|t_2-t_1|})}{(1-\rho^T)(1-\rho^2)},
$$

then

$$
E[(\tilde{X}_T(t_2) - \rho^{|t_2-t_1|}\tilde{X}_T(t_1))^2] = \frac{\sigma^2}{1-\rho^2}(1-\rho^{2|t_2-t_1|}).
$$

If in the last equation the "limit" $\rho \to 1$ is taken, the heuristic result is $E[(\tilde{X}_T(t_2) - \tilde{X}_T(t_1))^2] = \sigma^2 |t_2 - t_1|$, the defining condition for Brownian motion. But for T fixed, $\rho \to 1$, if and only if $\delta \to 0$, so that the limiting distributions as $\delta \to 0$ are formally related to a non-degenerate process, namely Brownian motion.

REFERENCES

GARABEDIAN, H. L. (1931): Relations between methods of summability. *Ann. Math. Stat.*, **32**, 83-106.

BIERENS DE HAAN, D. (1867): *Nouvelles Tables d' Integrales Definies*, Reprinted 1939, G. E. Stechert and Co., New York.

KOOPMANS, T. (1942): Serial correlation and quadratic forms in normal variables. *Ann. Math. Stat.*, **13**, 14-33.

LEHMANN, E. (1947): On optimum tests of composite hypotheses with one constraint. *Ann. Math. Stat.*, **18**, 473-494.

LEIPNIK, R. B. (1947): Distribution of the serial correlation coefficient in a circularly correlated universe. *Ann. Math. Stat.*, **18**, 80-87.

——— (1958a): Note on alternating Tannery series. *Am. Math. Monthly*, **65**, 3, 197-198.

——— (1958b): Moment generating functions of quadratic forms in serially correlated normal variables. *Biometrika*, **45**, 198-210.

MADOW, W. G. (1945): Note on the distribution of the serial correlation coefficient. *Ann. Math. Stat.*, **16**, 308-310.

MOORE, C. N. (1938): Summable series and convergence factors. *Amer. Math. Soc. Colloq. Publ.*, **22**, New York.

NEUMANN, J. V. (1942): Distribution of the ratio of the mean square successive difference to the variance. *Ann. Math. Stat.*, **12**, 367-395.

OGWARA, M. (1951): A note on the test of the serial correlation coefficient. *Ann. Math. Stat.*, **32**, 115-118.

QUENOUILLE, M. H. (1949): The joint distribution of serial correlation coefficients. *Ann. Math. Stat.*, **20**, 561-571.

SCHEFFE, H. (1947): A useful convergence theorem for probability distributions. *Ann. Math. Stat.*, **18**, 434-438.

WHITTAKER, J. M. and WATSON, G. N. (1948): *A Course of Modern Analysis*, Macmillan, London and New York.

STATISTICS PROPOSED FOR VARIOUS TESTS OF HYPOTHESES AND THEIR DISTRIBUTIONS IN PARTICULAR CASES*

By O. P. BAGAI

Punjab University, Chandigarh

SUMMARY. We find the distribution of the six statistics in multivariate analysis of variance. The distributions of the statistics for the cases $1 = 2, 3$ in the form of definite integrals and the limiting distributions of two the six statistics for the case $1 = 2, 3 \ldots 6$ in the form series, are also given.

1. INTRODUCTION

In multivariate analysis of variance (Pillai, 1954) the three tests of hypotheses, (I) equality of two dispersion matrices, (II) equality of the p-dimensional mean vectors and (III) the independence between a p-set and q-set of variates depend, when the respective hypotheses to be tested are true, only on the roots θ_i or ϕ_i $(i = 1, 2, \ldots, l)$ respectively of the determinantal equations

$$|A - \theta(A + C)| = 0 \qquad \ldots \quad (1.1)$$

and

$$|A - \phi C| = 0 \qquad \ldots \quad (1.2)$$

where A and C are independent sum of product (S.P.) matrices, based on sample observations with n_1 and n_2 degrees of freedom (d.f.) respectively and can be defined differently for different hypotheses.

The common standard form (Nanda, 1948; Roy, 1957) of the joint distribution of the eigenroots of (1.1) under the respective hypotheses, is as follows

$$C(m, n, l) \prod_{i=1}^{l} \theta_i^m (1 - \theta_i)^n \prod_{i=2}^{l} \prod_{j=1}^{i-1} (\theta_i - \theta_j) \prod_{i=1}^{l} d\theta_i \qquad \ldots \quad (1.3)$$

for

$$0 \leqslant \theta_1 \leqslant \theta_2 \leqslant \ldots \leqslant \theta_l \leqslant 1, \quad l = \min(p, n_1)$$

and

$$C(m, n, l) = \frac{\pi^{l/2} \prod_{i=1}^{l} \Gamma\left(\frac{2m + 2n + l + i + 1}{2}\right)}{\prod_{i=1}^{l} \Gamma\left(\frac{2m + i + 1}{2}\right) \Gamma\left(\frac{2n + i + 1}{2}\right) \Gamma\left(\frac{i}{2}\right)} \qquad \ldots \quad (1.4)$$

and the values of l, m, n for the respective hypotheses are as follows

(I) $\qquad l = p, \quad m = \frac{1}{2}(n_1 - p - 1), \quad n = \frac{1}{2}(n_2 - p - 1)$ $\qquad \ldots \quad (1.5)$

(II) If $p \leqslant n_1, \; l = p,$ $\qquad$ then $m = \frac{1}{2}(n_1 - p - 1), \; n = \frac{1}{2}(n_2 - p - 1)$

and if $p > n_1, \; l = n_1,$ $\qquad$ then $m = \frac{1}{2}(p - n_1 - 1), \; n = \frac{1}{2}(n_2 - p - 1)$ $\quad \ldots \quad (1.6)$

(III) *Same as* (II). $\qquad \ldots \quad (1.7)$

The common standard form (Hsu, 1939) of the joint distribution of the eigenroots of (1.2), under the respective hypotheses, is as follows

$$C(m, n, l) \prod_{i=1}^{l} \phi_i^m (1 + \phi_i)^{-(m+n+l+1)} \prod_{i=2}^{l} \prod_{j=1}^{i-1} (\phi_i - \phi_j) \prod_{i=1}^{l} d\phi_i \qquad \ldots \quad (1.8)$$

for

$$0 \leqslant \phi_1 \leqslant \phi_2 \leqslant \ldots, \leqslant \phi_l < \infty$$

where l, m, n and $C(m, n, l)$ are defined above as in (1.5), (1.6), (1.7) and (1.4) respectively.

Nanda (1948) gives the limiting form of (1.3) by setting $\theta_i = C_i/n$ and then letting $n \to \infty$. The limiting distribution is as follows

$$K(l, m) \prod_{i=1}^{l} C_i^m \exp\left[-\sum_{i=1}^{l} C_i \right] \prod_{i=2}^{l} \prod_{j=1}^{i-1} (C_i - C_j) \prod_{i=1}^{l} dC_i \qquad \ldots \quad (1.9)$$

where

$$K(l, m) = \pi^{l/2} \Big/ \prod_{i=1}^{l} \Gamma\left(\frac{2m+i+1}{2}\right) \Gamma\left(\frac{i}{2}\right) \qquad \ldots \quad (1.10)$$

and again l, m assume different values, defined above as in (1.5), (1.6) and (1.7) for the respective hypotheses.

2. Statistics proposed for tests of hypotheses (I), (II) and (III)

We list below the statistics based simultaneously on the roots of both the determinantal equations (1.1) and (1.2) which can be used to test the hypotheses (I), (II) and (III) with the suitable choice of the independent S.P., matrices A and C.

(i) Roy's statistics of largest, smallest and intermediate eigenroots based on the determinantal equation (1.1). We can simultaneously propose to include that of the eigenroots based on the determinantal equation (1.2).

(ii) Hotelling's T_k^2 statistic defined as follows :

$$T_k^2 = n_2 \operatorname{tr} (C^{-1}A) = n_2 \sum_{i=1}^{l} \left(\frac{\theta_i}{1-\theta_i} \right) = n_2 \sum_{i=1}^{l} (\phi_i).$$

(iii) Wilks' Λ statistic defined as follows :

$$\Lambda = \frac{|C|}{|A+C|} = \prod_{i=1}^{l} (1-\theta_i) = \prod_{i=1}^{l} (1+\theta_i)^{-1}.$$

(iv) The Wilks-Lawley U-statistic defined as follows :

$$U = |A| / |A+C| = \prod_{i=1}^{l} (\theta_i) = \prod_{i=1}^{l} \left(\frac{\phi_i}{1+\phi_i} \right).$$

(v) Pillai's V-statistic defined as follows :

$$V = \operatorname{tr} [(A+C)^{-1}A] = \sum_{i=1}^{l} (\theta_i) = \sum_{i=1}^{l} \left(\frac{\phi_i}{1+\phi_i} \right).$$

(vi) Finally, we propose another statistic Y defined as follows :

$$Y = \frac{|A|}{|C|} = \prod_{i=1}^{l} \left(\frac{\theta_i}{1-\theta_i} \right) = \prod_{i=1}^{l} \phi_i.$$

Of course, the distribution of any of the statistics, under the respective null hypotheses, can be found from either of the joint distributions (1.3) and (1.8), but it will be more convenient to use (1.3) for finding the distribution of Λ, U and V, (1.8) for that of T_k^2 and Y and either of the two for finding that of Roy's statistics.

Nanda (1948) gives the joint limiting form of (1.3), which we have listed under (1.1). Following him, the joint limiting form of (1.8) is easily proved also to be the same as (1.9) by setting $\phi_i = C_i/n$ in (1.8) and then letting $n \to \infty$. This fact that the joint limiting forms of both (1.3) and (1.8) are the same enables us to conclude that the limiting distributions of the statistics Y and U will be the same and also that of T_k^2 and V except for the constant multiplier. The same can be said in the case of Roy's statistics.

No great headway has been made so far in finding the distribution of the various statistics defined above. The classical T_k^2 is known (Rao, 1952) to be distributed, under the null hypotheses, as central chi-square with $n_1 p$ degrees of freedom (d.f.). In the case of non-centrality parameter $\tau_k^2 \neq 0$, the classical T_k^2 is a non-central chi-square distributed with $n_1 p$ d.f. The exact distribution of studentized T_k^2 for both central and non-central cases is not known in compact standard form. Ito (1956) has given, under the null hypotheses, its approximate distribution as an asymptotic expression of chi-squares each with $n_1 p$ d.f.

Wilks (1932) and Nair (1939) have given the exact distribution of Λ for $n = 1, 2$ and any p, and for $p = 1, 2$ and any n, by comparing the moments of Λ with those of F-ratio. Bartlett (1938) has suggested a useful approximation whose use has been made by Bartlett (1947) himself and by Rao (1952). More recently, Bannerjee (1958) has been able to give the exact distribution of Λ in series, but its tabular values are not yet available.

Roy (1943) and Nanda (1948a) have worked out the limiting and non-limiting distributions of the statistics—largest, smallest and intermediate eigenroots of the determinantal equation (1.1) to test the hypotheses (I), (II) and (III). Their tabular values have been given by Pillai (1957) for the case, $l = 2(1)5$, $m = 0(1)4$ and $n = 5$ to $1,000$ both at 5% and 1% significant levels.

Pillai (1954, 1955, 1957) has succeeded in giving an approximation to his statistic V and has been able to tabulate it for $l = 2(1)5$, $m = -.5(.5)5(5)80$, and $n = 5(5)80$. Nanda (1950) has also given the exact distribution for the special case when $m = 0$.

We have taken, in Section 3(a), the statistics T_k^2 and Y and have been able to give their distributions for $l = 2, 3$ in the form of definite integrals. Since the procedure is quite similar for the remaining statistics, we are not including them here. For them the reader may see Bagai (1960).

Finally, in Section 4(a), we list some integrals (published elsewhere). In Section 4(b), we have first found a new form suitable for finding the limiting distributions, for $l = 2, 3, 4$, of the statistics U or Y the form of series and for $l = 5, 6$ in the form of double definite integrals. This method can further be extended to any value of l.

It may be noted here that we have worked out another method of integrations different from that of Nanda (1948), of finding the limiting distribution of Roy's statistics. This method works very well in the case especially of the smallest eigen values. This new method of integration has further been demonstrated (Bagai, 1960) by solving some particular cases, giving various values to m, for $l = 2, 3$, and 4. Since the problem has already been completed by Roy, Nanda, and Pillai, we are not including our method in this publication.

3. DISTRIBUTIONS OF VARIOUS STATISTICS FOR $l = 2, 3$

(a) *Distributions of T_k^2 and Y, Case I; $l = 2$* : The joint distribution of ϕ_1 and ϕ_2 from (1.8) is,

$$C(m, n, 2)(\phi_1, \phi_2)^m [(1+\phi_1)(1+\phi_2)]^{-m-n-3}(\phi_2-\phi_1)d\phi_1\, d\phi_2 \qquad \dots \ (3.1)$$

where
$$0 \leqslant \phi_1, \quad \phi_2 < \infty.$$

(i) *For Y-Statistic*, let $\quad \phi_1\phi_2 = u, \ (1+\phi_1)(1+\phi_2) = v \qquad \dots \ (3.2)$

so that
$$(\phi_2-\phi_1)d\phi_1\, d\phi_2 = du\, dv$$

this changes (3.1) in the following form

$$C(m, n, 2) \ u^m v^{-2(m+n+3)} du\, dv. \qquad \dots \ (3.3)$$

411

Now the roots ϕ_1, ϕ_2 of the quadratic

$$x^2-(v-u-1)x+u=0 \qquad \dots (3.4)$$

are real if

$$(v-u-1)^2 \geqslant 4u$$

i.e., if

$$(1+\sqrt{u})^2 \leqslant v.$$

Then the limits for v and u are given by

$$(1+\sqrt{u}) \leqslant v < \infty$$

and

$$0 \leqslant u < \infty.$$

The distribution of $u(=\phi_1\phi_2)$ or Y is given by,

$$C(m, n, 2)u^m du \int_{v=(1+\sqrt{u})^2}^{\infty} v^{-m-n-3} dv \qquad \dots (3.5)$$

where

$$0 \leqslant u < \infty ;$$

or by

$$\frac{2C(m, n, 2)}{m+n+2} \frac{(\sqrt{u})^{2m+1}}{(1+\sqrt{u})^{2(m+1+2)}} d(\sqrt{u}) \qquad \dots (3.6)$$

where

$$0 \leqslant u < \infty.$$

(ii) *For T_k^2-statistic*, consider now the change,

$$\phi_1+\phi_2 = u \quad \text{and} \quad \phi_1\phi_2 = v. \qquad \dots (3.7)$$

Proceeding as above, the joint distribution (3.1) takes the following form :

$$C(m, n, 2)v^m(1+u+v)^{-m-n-3}du\, dv \qquad \dots (3.8)$$

where

$$0 \leqslant v \leqslant \tfrac{1}{4} u^2 \text{ and } 0 \leqslant u < \infty.$$

Then the distribution of u is

$$C(m, n, 2) \int_{v=0}^{u^2/4} v^m(1+u+v)^{-m-n-3}\ du\, dv. \qquad \dots (3.9)$$

Setting $v = (1+u)V_0$, we get, in place of (3.9), the following distribution

$$C(m, n, 2)(1+u)^{-n-2} du \int_{V_0=0}^{u^2/4(1+u)} V_0^m(1+V_0)^{-m-n-3}\, dV_0 \qquad \dots (3.10)$$

where

$$0 \leqslant u < \infty.$$

Case II ; F or $l = 3$. The joint distribution of ϕ_1, ϕ_2, ϕ_3 from (1.8) is,

$$C(m, n, 3)(\phi_1\phi_2\phi_3)^m[(1+\phi_1)(1+\phi_2)(1+\phi_3)]^{-m-n-4} \prod_{i=2}^{3} \prod_{j=1}^{i-1} (\phi_i-\phi_j) \prod_{i=1}^{3} d\phi_i. \qquad \dots (3.11)$$

where

$$0 \leqslant \phi_1 \leqslant \phi_2 \leqslant \phi_3 < \infty.$$

For finding the distributions of both the statistics Y and T_k^2 for three $k = 3$ eigenroots, we effect the following changes

$$\phi_1+\phi_2+\phi_3 = u, \phi_1\phi_2+\phi_1\phi_3+\phi_2\phi_3 = v, \text{ and } \phi_1\phi_2\phi_3 = w \qquad \dots (3.12)$$

so that

$$(\phi_3-\phi_2)(\phi_2-\phi_1)(\phi_3-\phi_1)d\phi_1 d\phi_2 d\phi_3 = du\, dv\, dw.$$

Then the relation (3.11) reduces to

$$C(m, n, 3)w^m(1+u+v+w)^{-m-n-4}\, du\, dv\, dw \qquad \qquad \dots \quad (3.13)$$

where ϕ_1, ϕ_2, ϕ_3 are the roots of the cubic,

$$x^3-ux^2+vx-w=0. \qquad \dots \quad (3.14)$$

(i) *For Y-statistics*, in order that the roots of the cubic (3.14) are real and positive, we can (Bagai, 1961) write the limits on u, v and w as follows

$$0 \leqslant w < \infty \qquad \text{and} \qquad \theta \leqslant w < \infty$$

$$3w^{2/3} \leqslant v \leqslant 3w^{2/3}(1+\sqrt{3}) \qquad 3w^{2/3}(1+\sqrt{3}) \leqslant v < \infty, \qquad \qquad \dots \quad (3.15)$$

$$\beta_3 \leqslant u \leqslant \beta_4; \quad \beta'_3 \leqslant u \leqslant \beta_4$$

where $\qquad \beta_3 = \max\left[\sqrt{3v}, \dfrac{1}{4w}\left\{2(24vw^2+\dfrac{1}{9}v^4)^{\frac{1}{2}}\cos\dfrac{\phi}{3}+\dfrac{v^2}{3}\right\}\right]$,

$$\beta_4 = \dfrac{1}{4w}\left[-2\left(24vw^2+\dfrac{v^4}{9}\right)^{\frac{1}{2}}\cos\dfrac{\pi+\phi}{3}+\dfrac{v^2}{3}\right]$$

$$\beta_3' = \max\left[\sqrt{3v}, \dfrac{1}{4w}\left[-2\left(24vw^2+\dfrac{v^4}{9}\right)^{\frac{1}{2}}\cos\dfrac{\phi_1+\pi}{3}+\dfrac{v^2}{3}\right]\right]$$

and $\qquad \beta_4' = \dfrac{1}{4w}\left[2\left(24vw^2+\dfrac{v^4}{9}\right)^{\frac{1}{2}}\cos\dfrac{\phi_1}{3}+\dfrac{v^2}{3}\right] \qquad \dots \quad (3.16)$

where ϕ_1 is the supplement of ϕ, and ϕ is defined by,

$$\tan\phi = \dfrac{-8\left(\dfrac{1}{3}v^3-9w^2\right)^{3/2}}{2\left(216w^4+20v^3w^2-\dfrac{1}{27}v^6\right)} \qquad \dots \quad (3.17)$$

for $\qquad \left(216w^4+20v^3w^2-\dfrac{v^6}{27}\right)$ being positive.

Thus, the distribution of $w(=\phi_1\phi_2\phi_3)$ or Y for 3 eigenroots from (3.13) is the following

$$C(m, n, 3)\, w^m \int_v \int_u (1+u+v+w)^{-m-n-4}\, du\, dv\, dw \qquad \dots \quad (3.18)$$

where u, v and w are defined in (3.15).

Effecting the following change in (3.18)

$$v = (1+w)V_1, \quad u = (1+w)(1+V_1)u_1, \qquad \dots \quad (3.19)$$

so that $du\, dv = (1+w)^2(1+V_1)dV_1\, dU_1$, we get in place of (3.18), the following

$$C(m, n, 3)\dfrac{w^m}{(1+w)^{m+n+3}}\int_{V_1}\int_{U_1}\dfrac{dV_1 dU_1}{(1+V_1)^{m+n+3}(1+U_1)^{m+n+4}} \qquad \dots \quad (3.20)$$

for $\qquad 0 \leqslant w < \infty \quad \text{and} \quad 0 \leqslant w < \infty.$

$$\dfrac{3w^{2/3}}{1+w} \leqslant V_1 < \dfrac{3w^{2/3}(1+\sqrt{3})}{1+w} \quad \dfrac{3(1+\sqrt{3})w^{2/3}}{1+w} \leqslant V_1 < \infty,$$

$$\dfrac{\beta_3}{(1+w)(1+V_1)} \leqslant U_1 \leqslant \dfrac{\beta_4}{(1+w)(1+V_1)} \quad \dfrac{\beta_3'}{(1+w)(1+V_1)} \leqslant U_1 \leqslant \dfrac{\beta_4}{(1+w)(1+V_1)}$$

where β_2, β_4, β'_2 and β'_4 are defined as in (3.16), and u used in them is equal to $(1+w)V_1$.

(ii) *For T_k^2-statistic*, in order that the roots of the cubic (3.14) be real and positive, we write down the limits (Bagai, 1961) respectively for u, v, w as follows

$$0 \leqslant u < \infty \quad \text{and} \quad 0 \leqslant u < \infty,$$

$$0 \leqslant v \leqslant \tfrac{1}{4} u^2 \quad \text{and} \quad \tfrac{1}{4} u^2 \leqslant v \leqslant \tfrac{1}{3} u^2,$$

$$0 \leqslant w \leqslant \beta_2 \quad \text{and} \quad \beta_1 \leqslant w \leqslant \beta_2, \qquad \ldots \quad (3.21)$$

where β_1 and β_2 are defined as in (3.26) below.

Thus, the distribution of $u(= \phi_1 + \phi_2 + \phi_3)$ or T_k^2 for $k = 3$.

Eigenroots, from (3.13) are the following

$$C(m, n, 3) \int\limits_{0} \int\limits_{w} w^m (1+u+v+w)^{-m-n-4} \, du \, dv \, dw \qquad \ldots \quad (3.22)$$

with limits for u, v and w shown in (3.21)

Effecting another change in (3.22) as follows

$$v = (1+u) V_2, \quad w = (1+u)(1+V_2) U_2 \qquad \ldots \quad (3.23)$$

we get the distribution of $u = T_k^2$ for $k = 3$ eigenroots from (3.22) as follows

$$C(m, n, 3) \, \frac{du}{(1+u)^{n+2}} \int\limits_{V_2} \int\limits_{U_2} \frac{U_2^m \, dV_2 \, dU_2}{(1+V_2)^{n+3} (1+U_2)^{m+n+4}} \qquad \ldots \quad (3.24)$$

where
$$0 \leqslant u < \infty \quad \text{and} \quad 0 \leqslant u < \infty,$$

$$0 \leqslant V_2 \leqslant \frac{u^2}{4(1+u)} \quad \text{and} \quad \frac{u^2}{4(1+u)} \leqslant V_2 \leqslant \frac{u^2}{3(1+u)},$$

$$0 \leqslant U_2 \leqslant \frac{\beta_2}{(1+u)(1+V_2)} \qquad \frac{\beta_1}{(1+u)(1+V_2)} \leqslant U_2 \leqslant \frac{\beta_2}{(1+u)(1+V_2)} \quad \ldots \quad (3.25)$$

where β_1 and β_2 are respectively defined below as

$$\beta_1 = \frac{1}{3} u \left(v - \frac{2}{9} u^2 \right) - \frac{2}{27} (u^2 - 3v)^{3/2} \text{ and } \beta_2 = \frac{1}{3} u \left(v - \frac{2}{9} u^2 \right) + \frac{2}{27} (u^2 - 3v)^{3/2}$$

$$\ldots \quad (3.26)$$

and v used in them is equal to $(1+u)V_2$.

4. Limiting distributions of U or Y for $l = 2(1)6$

(a) *Certain integrals.* We first list the following integrals which are made use of for finding the limiting distributions of U or Y.

(i) We make use of the Legendre's duplication formula for the gamma function, namely of

$$\Gamma(n+\tfrac{1}{2}) \dot\Gamma(n+1) = \frac{\sqrt{\pi} . \Gamma.(2n+1)}{2^{2n}}. \qquad \ldots \quad (4.1)$$

(ii) A definite integral (Larson, 1948), namely,

$$\int\limits_{0} \exp [-x^2 - ax^{-2}] dx = \frac{\sqrt{\pi}}{2} \exp (-2a) \qquad \ldots \quad (4.2)$$

for $a > 0$, is made use of quite frequently.

(iii)   Now, we give the following two definite integrals (Bagai, 1960) which we have evaluated ourselves for the purpose : consider

$$\text{(a)} \quad I = 4 \int_0^\infty x \exp\left[-2(x+ax^{-1})\right]dx,$$

then

$$I = [1+2\gamma-\log\ 4a]\left[\frac{(4a)^2}{2!0!}+\frac{(4a)^3}{3!1!}+\frac{(4a)^4}{4!2!}+\ldots\right]$$

$$+\left[1-4a+\frac{(4a)^2}{2!0!}\cdot\tfrac{1}{2}+\frac{(4a)^3}{3!1!}(\tfrac{1}{3}+\tfrac{1}{2}+1)+\frac{(4a)^4}{4!2!}(\tfrac{1}{4}+\tfrac{1}{3}+\tfrac{1}{2}+1+\tfrac{1}{2})+\frac{(4a)^5}{5!3!}(\tfrac{1}{5}+\ldots+\tfrac{1}{2}+1+\tfrac{1}{2}+\tfrac{1}{3})+\ldots\right]$$

$$\ldots \quad (4.3)$$

where γ is the Euler's constant.

$$\text{(b)} \quad L(a) = 2 \int_0^\infty x \exp(-x^2-ax^{-1})dx$$

$$L(a) = [\gamma-\log a]\left[2\cdot\frac{a^2}{2!}-\frac{2^2}{2}\frac{a^4}{4!}+\frac{2^3}{2.4}\cdot\frac{a^6}{6!}-\ldots\right]$$

$$+\left[1+\frac{2a^2}{2!}(\tfrac{1}{2}+1)-\frac{2^2a^4}{4!}\cdot\tfrac{1}{2}(\tfrac{1}{4}+\tfrac{1}{3}+\tfrac{1}{2}+1+\tfrac{1}{2})+\frac{2^3a^6}{6!}\ \frac{1}{2.4}(\tfrac{1}{6}+\tfrac{1}{5}+\ldots+\tfrac{1}{2}+1+\tfrac{1}{2}+\tfrac{1}{4})-\ldots\right]$$

$$-\sqrt{\pi}\left[a-\frac{2}{1}\cdot\frac{a^3}{3!}+\frac{2^2}{1.3}\cdot\frac{a^5}{5!}-\frac{2^3}{1.3.5}\frac{a^7}{7!}+\ldots\right] \qquad \ldots \quad (4.4)$$

where again γ is the Euler's constant.

It may be noted that in place of definite integral (4.3), we can also, by letting $r = 2$, $p = 2$ and $L = 8a$, use the following integral [Bateman, pp. 146(29)] giving the use of modified Bessel's function whose tabular values are readily available, i.e. for both p and L as positives,

$$\left.\begin{array}{l}
\displaystyle\int_0^\infty t^{r-1}\exp\left(-pt-\frac{L}{4t}\right)dt = 2\left(\frac{L}{4p}\right)^{\frac{1}{2}r}K_r(\sqrt{Lp}) \\[3mm]
\text{where} \qquad K_r(y) = \dfrac{\pi}{2}\cdot\dfrac{I_{-r}(y)-I_r(y)}{\sin \pi y} \\[3mm]
\text{and} \qquad I_r(y) = \displaystyle\sum_{m=0}^\infty\left[\dfrac{(\tfrac{1}{2}y)^{r+2L}}{m!\Gamma(r+m+1)}\right]
\end{array}\right\} \qquad \ldots \quad (4.5)$$

Distributions.   The moment generating function of (1.9) is,

$$m(t) = \int_0^\infty\ldots\int_0^\infty K(l, m)(C_1C_2\ldots C_l)^m\exp\left[-\sum_{i=1}^l\ C_i+t\sum_{i=1}^l C_i\right]\prod_{i=2}^l\prod_{j=1}^{i-1}(C_i-C_j)\prod_{i=1}^l dC.$$

from which the h-th moment μ_h' about the origin is,

$$\mu_h' = \frac{K(l, m)}{K(l, m+h)} = \prod_{i=1}^l\cdot\frac{\Gamma\left(\frac{2m+i+1}{2}+h\right)}{\Gamma\left(\frac{2m+i+1}{2}\right)}.$$

This h-th moment shows that the moments of the limiting distribution of the product of the roots $(C_1 C_2 \ldots C_l)$ can be determined from the following

$$\frac{\exp(-v_1)}{\Gamma\left(\frac{2m+2}{2}\right)} v_1^{(2m+2/2)-1} dv_1 \cdot \frac{\exp(-v_2)}{\Gamma\left(\frac{2m+3}{2}\right)} v_2^{(2m+3/2)-1} dv_2 \cdots \frac{\exp(-v_l)}{\Gamma\left(\frac{2m+l+1}{2}\right)} v_l^{(2m+l+1/2)-1} dv_l$$

or from

$$\frac{\exp\left(-\sum_{i=1}^{l} v_i\right)}{\Gamma(m+1)\Gamma\left(m+\frac{3}{2}\right)\cdots\Gamma\left(m+\frac{l+1}{2}\right)} v_1^m v_2^{m+\frac{1}{2}} \cdots v_l^{m+(l-1/2)} \prod_{i=1}^{i} dv_i \qquad \ldots \quad (4.6)$$

where
$$0 \leqslant v_i < \infty, \quad i = 1, 2, \ldots, l.$$

Now, we determine the limiting distribution of the product $(C_1 C_2 \ldots C_l)$ of the roots for particular values of l. For $l = 2$, the result (Rao, 1952) is already known. We have also (Bagai, 1960) determined independently that $2\sqrt{C_1 C_2}$ is a gamma variate of parameter $(2m+1)$.

For $l = 3$, substituting $l = 3$ in (4.6), we get the joint distribution of v_1, v_2, and v_3 as follows

$$\frac{\exp(-v_1-v_2-v_3)}{\Gamma(m+1)\Gamma\left(m+\frac{3}{2}\right)\Gamma(m+2)} v_1^m \cdot v_2^{m+\frac{1}{2}} \cdot v_3^{m+1} dv_1 dv_2 dv_3 \qquad \ldots \quad (4.7)$$

where
$$0 \leqslant v_1, v_2, v_3 < \infty.$$

Making use of (4.1), the joint distribution (4.7) takes the following form

$$\frac{2^{2m+2}}{\sqrt{\pi}} \cdot \frac{\exp(-v_1-v_2-v_3)}{\Gamma(m+1)\Gamma(2m+3)} v_1^m \cdot v_2^{m+\frac{1}{2}} \cdot v_3^{m+1} dv_1 dv_2 dv_3. \qquad \ldots \quad (4.8)$$

Setting $4v_1 v_2 v_3 = V_1^2$, $v_2 v_3^2 = V_2^2$ and $v_3 = V_3^2$, (4.8) reduces to

$$\frac{2^3}{\sqrt{\pi}} \cdot \frac{V_1^{2m+1} \cdot V_3^{-1}}{\Gamma(m+1)\Gamma(2m+3)} \exp\left(-\frac{V_1^2 V_3^2}{4 V_2^2} - \frac{V_2^2}{V_3^2} - V_3^2\right) dV_1 dV_2 dV_3 \qquad \ldots \quad (4.9)$$

where
$$0 \leqslant V_1, V_2, V_3 < \infty.$$

Integrating (4.9) with respect to V_2 with the help of (4.2), we obtain the joint distribution of V_1, V_3 as follows

$$\frac{4 V_1^{2m+1} V_3}{\Gamma(m+1)\Gamma(2m+3)} \exp\left(-\frac{V_1}{V_3} - V_3^2\right) dV_1 dV_3 \qquad \ldots \quad (4.10)$$

where
$$0 \leqslant V_1, V_3 < \infty.$$

Again integrating (4.10) with respect to V_3, we get the distribution of $V_1 = 2\sqrt{v_1 v_2 v_3}$ (or $= 2\sqrt{C_1 C_2 C_3}$) as follows

$$\frac{2 V_1^{2m+1}}{\Gamma(m+1)\Gamma(2m+3)} L(V_1) dV_1 \qquad \ldots \quad (4.11)$$

where $0 \leqslant V_1 < \infty$ and $L(V_1)$ is defined in (4.4) by setting $a = V_1$.

For $l = 4$, substituting $l = 4$ in (4.6), we get the joint distribution of v_1, v_2, v_3 and v_4 as follows

$$\frac{\exp\left(-\sum_{i=1}^{4} v_i\right)}{\Gamma(m+1)\Gamma\left(m+\frac{3}{2}\right)\Gamma(m+2)\Gamma\left(m+\frac{5}{2}\right)} v_1^m \cdot v_2^{m+\frac{1}{2}} \cdot v_3^{m+1} \cdot v_4^{m+3/2} \prod_{i=1}^{4} dv_i \qquad \ldots \quad (4.12)$$

where
$$0 \leqslant v_1, v_2, v_3, v_4 < \infty.$$

Making use of (4.1), the joint distribution (4.12) reduces to,

$$\frac{2^{4m+4}}{\pi} \frac{\exp\left(-\sum\limits_{i=1}^{4} v_i\right)}{\Gamma(2m+2)\Gamma(2m+4)} \, v_1^m \cdot v_2^{m+\frac{1}{2}} \cdot v_3^{m+1} \cdot v_4^{m+3/2} \prod_{i=1}^{4} dv_i \qquad \ldots \ (4.13)$$

where

$$0 \leqslant v_1, \ v_2, \ v_3, \ v_4 < \infty.$$

Setting $16v_1v_2v_3v_4 = V_1^2$, $v_2v_3v_4 = V_2^2$, $v_3v_4 = V_3^2$ and $v_4 = V_4^2$ in (4.13), we get the joint distribution of V_1, V_2, V_3 and V_4 as follows

$$\frac{16}{\pi} \frac{V_1^{2m+1}}{\Gamma(2m+2)\,\Gamma(2m+4)} \exp\left(-\frac{V_1^2}{16\,V_2^2} - \frac{V_2^2}{V_3^2} - \frac{V_3^2}{V_4^2} - V_4^2\right) dV_i \qquad \ldots \ (4.14)$$

where

$$0 \leqslant V_1, V_2, \ V_3 \ \text{and} \ V_4 < \infty.$$

With the help of (4.2) we integrate (4.14) successively with respect to V_2 and V_4 and obtain the joint distribution of V_1 and V_3 as follows

$$\frac{4 V_1^{2m+1}}{\Gamma(2m+2)\,\Gamma(2m+4)} \ V_3 \exp\left(-\frac{2V_1}{4 V_3} - 2V_3\right) dV_1 dV_3 \qquad \ldots \ (4.15)$$

where

$$0 \leqslant V_1, \ V_3 < \infty.$$

Now, making use of either (4.3) or (4.5), we can integrate out V_3 from (4.15) to obtain the distribution of V_1. With the use of (4.3) we obtain the distribution of $V_1(= 4\sqrt{v_1 v_2 v_3 v_4})$ or $4\sqrt{C_1 C_2 C_3 C_4}$ as follows

$$\frac{V_1^{2m+1} dV_1}{\Gamma(2m+2)\,\Gamma(2m+4)} \left\{ \left[(1+2\gamma) - \log V_1\right] \left[\frac{V_1^2}{2!0!} + \frac{V_1^3}{3!1!} + \frac{V_1^4}{4!2!} + \ldots \ldots \right] \right.$$

$$+ \left[1 - V_1 + \frac{V_1^2}{2!0!} \cdot \frac{1}{2} + \frac{V_1^3}{3!1!} \left(\frac{1}{3} + \frac{1}{2} + 1\right) + \frac{V_1^4}{4!2!} \left(\frac{1}{4} + \frac{1}{3} + \frac{1}{2} + 1 + \frac{1}{2}\right) + \frac{V_1^5}{5!3!} \right.$$

$$\left. \left. \left(\frac{1}{5} + \ldots + \frac{1}{2} + 1 + \frac{1}{3} + \frac{1}{5}\right) + \ldots \right] \right\} \qquad \ldots \ (4.16)$$

where $0 \leqslant V_1 < \infty$, and γ is the Euler's constant. Or with the use of (4.5), we obtain the alternative form of the distribution of $V_1(= 4\sqrt{v_1 v_2 v_3 v_4})$ or $4\sqrt{C_1 C_2 C_3 C_4}$ as follows

$$\frac{2 V_1^{2m+1}}{\Gamma(2m+2)\,\Gamma(2m+4)} \ K_2(2\sqrt{V_1}) dV_1 \qquad \ldots \ (4.17)$$

where $0 \leqslant V_1 < \infty$ and $K_2(2\sqrt{V_1})$ is obtained as defined in (4.5).

For $l = 5, 6$: A similar method, as used above for $l = 3, 4$, was applied for the cases $l = 5, 6$ and the following distributions in the form of definite integrals were obtained.

(i) The distribution of $V_1(= C_1 C_2 C_3 C_4 C_5)$ is as follows

$$\frac{2^{4m+8} \cdot V_1^m dV_1}{\Gamma(m+1)\,\Gamma(2m+3)\,\Gamma(2m+5)} \int\limits_{V_2=0}^{\infty} \int\limits_{V_4=0}^{\infty} V_4 \exp\left(-\frac{V_1}{V_2^2} - \frac{2V_2}{V_4} - 2V_4\right) dV_2 dV_4 \qquad \ldots \ (4.18)$$

where

$$0 \leqslant V_1 < \infty.$$

(ii) The distribution for $V_1(= C_1 C_2 C_3 C_4 C_5)$ is as follows

$$\frac{2^{6m+1} \cdot V_1^m dV_1}{\Gamma(2m+2)\,\Gamma(2m+4)\,\Gamma(2m+6)} \int\limits_{V_5=0}^{\infty} \int\limits_{V_6=0}^{\infty} \exp\left(-\frac{2V_1}{V_3} - \frac{2V_3}{V_5} - 2V_5\right) dV_3 dV_5 \qquad \ldots \ (4.19)$$

where

$$0 \leqslant V_1 < \infty.$$

References

Bagai, O. P. (1960) : *Ph.D. Thesis.* University of British Columbia, Vancouver, Canada.

————— (1961) : Bounds for the coefficients of a certain cubic. *Mathematics Student India.*

Banerjee, D. P. (1958) : On the exact distribution of a test in multivariate analysis. *J. Roy. Stat. Soc.,* B, **20,** 108-110.

Bartlett, M. S. (1938) : Further aspects of the theory of multiple regression. *Proc. Camb. Phil. Soc.,* **34,** 33-40.

————— (1947) : Multivariate analysis. *Supp. J. Roy. Stat. Soc.,* **9,** 176-196.

Hotelling, H. (1950) : A generalized T test and measure of multivariate dispersion. *Proc. Second Berkeley Symposium On Math. Stat. and Prob.,* 23-41. Berkeley, University of California Press.

Hsu, P. L. (1939): On the distribution of roots of certain determinatal equations. *Ann. Eugen.,* **9,** 250-258.

Ito, K. (1956) : Asymptotic formula for the distribution of Hotelling's generalized T-statistic. *Ann. Math. Stat.,* **27,** 1091-1105.

Larson, H. D. (1948) : *Rinehart Mathematical Tables, Formulas and Curves,* New York, Rinehart and Co. Inc. publishers.

Lawley, D. N. (1938) : Tests of significance for the latent roots of covariance and correlation matrice. *Biometrika,* **43,** 128-136.

Nair, U. S. (1939) : The application of moment function in the study of distribution laws in statistics. *Biometrika,* **3,** 274-294.

Nanda, D. N. (1948a) : Distribution of a root of a determinatal equation. *Ann. Math. Stat.,* 10, 47-57.

————— (1948b) : Limiting distributions of a root of a determinantal equation. *Ann. Math. Stat.,* **19,** 340-350.

————— (1950) : Distribution of the sum of roots of a determinantal equation under a certain condition. *Ann. Math. Stat.,* **21,** 432-439.

Pillai, K. C. S. (1954) : On the distribution of Hotelling's generalized T test. Abstract, *Ann. Math. Stat.,* **25,** 412.

————— (1955) : Some new criteria in multivariate analysis. *Ann. Math. Stat.,* **26,** 117-121.

————— (1957) : *Concise Tables for Statistics,* Manila, The Statistical Centre, University of Phillipines.

Rao, C. R. (1952) : *Advanced Statistical Methods in Biometric Research,* John Wiley and Sons, New York.

Roy, S. N. (1939) : p-statistics on more generalization in analysis of variance appropriate to multivariate problems. *Sankhyā,* **4,** 381-396.

————— (1943) : The sampling distribution of p-statistic and certain allied statistics on the non-null hypothesis. *Sankhyā,* **6,** 15-34.

————— (1945) : The individual sampling distribution of the maximum, the minimum and any intermediate of the p-statistics on the null hypothesis. *Sankhyā,* **7,** 133-150.

————— (1957) : *Some Aspects of Multivariate Analysis,* John Wiley and Sons, New York.

Wilks, S. S. (1932) : Certain generalizations in the analysis of variance. *Biometrika,* **26,** 471-494.

ENTROPY OF LOGARITHMIC SERIES DISTRIBUTIONS

By GIFT SIROMONEY

Madras Christian College, Tambaram

SUMMARY. The logarithmic series distribution was introduced by Fisher and Williams. We show that there are two types of logarithmic series distributions, which we shall call Type I and Type II. We evaluate the entropies of the two distributions and present the corresponding interpretations.

Two types of logarithmic series distributions. The logarithmic series distribution was introduced first by Fisher (1943) in connection with some biological populations. Williams (1944, 1947) continued to present many other data which conform to the logarithmic series type. We now show that there are two kinds of logarithmic series distributions.

Type I. Consider a sample of flea-ridden mice, classified according to the number of fleas on each. This is an example of a distribution over the integers 1, 2, ... with relative frequencies p_1, p_2, ... with $p_r = \beta^r/[-ln(1-\beta)]r$. We choose to call this distribution Type I logarithmic series distribution. Other examples of the distribution include the distributions of wet and dry spells. (Williams, 1952 ; Ramabhadran, 1954).

Let H_1 be the entropy of this distribution.

Then
$$H_1 = \sum_{r=1}^{\infty} p_r \, ld \, p_r$$

where
$$p_r = \frac{\text{number of mice with } r \text{ fleas}}{\text{total number of mice}}$$

$$= \frac{\beta^r}{r\{-ln(1-\beta)\}} \text{ and is small for large } r$$

and
$$0 < \beta < 1.$$

H_1 measures the average amount of uncertainty per mouse with reference to the number of fleas on a mouse and only flea-ridden mice are considered.

$$H_1 = -\sum_1^{\infty} \left[\frac{\beta^r ld\,\beta}{\{-ln(1-\beta)\}} - \frac{\beta^r ld\,r}{\{-ln(1-\beta)\}} - \frac{\beta^r ld\{-ln(1-\beta)\}}{r\{-ln(1-\beta)\}} \right]$$

$$= \frac{ld\beta}{ln(1-\beta)} \left(\frac{\beta}{1-\beta} \right) + ld\{-ln(1-\beta)\} + \frac{1}{\{-ln(1-\beta)\}} \sum_1^{\infty} \frac{(ld\,r)\beta^r}{r} \; bits,$$

$\Sigma \frac{ld\,r}{r} \beta^r$ is uniformly convergent for $0 < \delta_1 \leqslant \beta \leqslant \delta_2 < 1$ where δ_1 and δ_2 are arbitrary. $\frac{ld\,\beta}{ln(1-\beta)} \cdot \left(\frac{\beta}{1-\beta} \right)$, $ld\{-ln(1-\beta)\}$ and $\frac{1}{-ln(1-\beta)} \sum_1^{\infty} \frac{(ld\,r)\beta^r}{r}$ are all monotonic functions of β. Hence H_1 is a monotonic increasing function of β.

Thus β itself is a measure of the uncertainty associated with the number of fleas on a mouse from the sample and we shall call β, the index of uncertainty.

Type II. Let f_r represent the frequency of species of exactly r butterflies (or birds or insects) each, in a population of butterflies of a given locality classified according to the species. The Type II logarithmic series distribution equates f_r to $\dfrac{\alpha\beta^r}{r}$, $r = 1, 2\ldots$. It should be noted that this is different from what was done in the case of Type I distribution. It is a distribution on the set of all species. Thus, if a butterfly caught belongs to a species with exactly r butterflies, its probability is

$$r/\text{total number of butterflies} = \frac{r}{\Sigma r f_r} = \frac{r(1-\beta)}{\alpha\beta},$$

and there are $\dfrac{\alpha\beta^r}{r}$ species with exactly r butterflies. Let H_2 denote the entropy in this case. Then H_2 measures the average amount of uncertainty per butterfly with reference to the species.

$$H_2 = -\overset{\alpha\beta}{\underset{1}{\Sigma}} \frac{(1-\beta)}{\alpha\beta} \; ld \left(\frac{1-\beta}{\alpha\beta} \right) - \overset{\alpha\beta^2/2}{\underset{1}{\Sigma}} \frac{2(1-\beta)}{\alpha\beta} \; ld \left(\frac{2(1-\beta)}{\alpha\beta} \right) - \ldots$$

$$= -\alpha\beta \left(\frac{1-\beta}{\alpha\beta} \right) ld \left(\frac{1-\beta}{\alpha\beta} \right) - \frac{\alpha\beta^2}{2} \left(\frac{2(1-\beta)}{\alpha\beta} \right) ld \left(\frac{2(1-\beta)}{\alpha\beta} \right) - \ldots$$

$$= ld \, \alpha + ld \, \frac{\beta}{1-\beta} - (1-\beta) \overset{\infty}{\underset{1}{\Sigma}} \beta^r ld. \, (1+r) \; bits,$$

$\overset{\infty}{\underset{1}{\Sigma}} \beta^r ld \,(1+r)$ is absolutely convergent and $0 < \beta < 1$. When α is fixed, H_2 increases with β.

Let T_1 and T_2 be the number of butterflies in two samples from the same population. Then $T_1 = \dfrac{\alpha\beta_1}{1-\beta_1}$ and $T_2 = \dfrac{\alpha\beta_2}{1-\beta_2}$. If $T_1 < T_2$ then $\beta_1 < \beta_2$. Therefore H_2 is greater for larger samples from the same biological population.

I wish to thank Dr. W. F. Kibble of the Madras Christian College for his guidance.

References

CARLESON, L. (1958) : Two remarks on the basic theorems on information theory. *Mathematica Scandinavica*, **6**, 175-80.

FISHER, R. A., CORBET, A. S. and WILLIAMS, C. B. (1943) : Species and the number of individuals in a random sample of an animal population. *Journal of Animal Ecology*, **12**, 42.

SHANNON, C. E. (1948) : A mathematical theory of communication. *Bell System Technical Journal*, **27**, 379-423.

RAMABHADRAN, V. K. (1954) : A statistical study of the persistency of rain days during the monsoon season at Poona. *Indian Journal of Meteorology and Geophysics*, **5**, 48-55.

WILLIAMS, C. B. (1944) : Some applications of the log series and the index of diversity to ecological problems. *Journal of Ecology*, **32**, 1-44.

———— (1947) : The logarithmic series and its applications to biological problems. *Journal of Ecology*, **34**, 253-272.

———— (1952) : Sequence of wet and dry days considered in relation to the logarithmic series. *Quarterly Journal of the Royal Meteorological Society*, **78**, 91-96.

SOME SAMPLING SCHEMES IN PROBABILITY SAMPLING

By T. V. HANURAV

Indian Statistical Institute

SUMMARY. In this paper we give a sampling scheme for finite populations, resulting in a given set of inclusion probabilities. Also, we give another scheme with the property that it can be stopped at any stage without distorting the proportionate values of the inclusion probabilities.

1. INTRODUCTION

We consider a finite population consisting of N units

$$u_1, u_2, ..., u_N. \qquad \qquad ... \ (1.1)$$

A sample 's' from (1.1) is defined as an ordered set of units from (1.1). A sampling design is a set 'S' of samples 's' from (1.1), with a probability measure 'P' defined on it, and is denoted by

$$D = D(S, P). \qquad \qquad ... \ (1.2)$$

For any unit u_i of (1.1), let

$$\pi_i = \sum_{s \supset u_i} P(s) \quad \text{for} \quad 1 \leqslant i \leqslant N \qquad \qquad ... \ (1.3)$$

where the summation extends over all samples 's' of (1.2) which contain u_i at least once. Similarly, for any pair u_i and u_j, let

$$\pi_{ij} = \sum_{s \supset u_i, u_j} P(s), \quad \text{for} \quad \begin{matrix} 1 \leqslant i, j < N. \\ i \neq j \end{matrix} \qquad \qquad .... \ (1.4)$$

π_i is called the inclusion probability of u_i and π_{ij} the joint inclusion probability of the pair u_i and u_j. Clearly, these are the probabilities that a sample 's' from (1.2) contains the unit u_i, the pair of units u_i and u_j, respectively.

When we have to estimate the total of a character Y for the entire population (1.1), we have practical situations where the values of another character X, which is closely proportional to Y, are available for each unit of (1.1). It is known that in such situations, the selection of a suitable sampling design which takes into account these values of X results in obtaining estimators of the total of Y which are better than those in which this information is not incorporated. A commonly advocated procedure is to select the units from (1.1) with probabilities proportional to the corresponding value of X, with or without replacement, or systematically. These are known as pps sampling procedures. However, it was felt (Horvitz and Thompson, 1952) that the selection of the units such that the resulting π_i's are proportional to the corresponding X values, may even be better in some cases. The problem that arises now is the construction of a sampling design (or equivalently a sampling scheme, as pointed out by the author (1962)) which results in a given set of π_i's.

This problem was first posed by Horvitz and Thompson (1952) where they attempted a solution. However, they limited their attention to the special case when each sample contains only two distinct units, and even in that case the method that they gave has rather severe limitations which they themselves pointed out. Further, the calculation of π_i's from the probabilities of selection at each draw, is quite complicated when the samples are drawn without replacement (as they did) even when the sample size is small, as can be seen from their paper as also from the paper by Yates and Grundy (1953).

In practice, it is necessary that we should not only be able to give an estimator of the population total of Y, say $\hat{Y}$ but also an unbiased estimator of its variance. For the estimators known to us, it is necessary that

$$\pi_i > 0 \quad \text{and} \quad \pi_{ij} > 0 \quad \text{for all} \quad 1 \leqslant i, j \leqslant N.$$
$$i \neq j$$

Hence we require a sampling scheme which while resulting in a given set of π_i's, is suitable for an easy calculation of π_{ij}'s. Certainly, a scheme consisting of independent draws is advantageous for this purpose. For, if the draws are independent and $p_i^{(r)}$ is the probability of selecting u_i in the r-th draw $(1 \leqslant i \leqslant N)$, we have

$$\pi_{ij} = 1 - \prod_{r=1}^{n} \{1 - p_i^{(r)}\} - \prod_{r=1}^{n} \{1 - p_j^{(r)}\} + \prod_{r=1}^{n} \{1 - p_i^{(r)} - p_j^{(r)}\} \qquad \ldots \quad (1.5)$$

where n is the total number of draws made.

In Section 2, we give a sampling scheme which results in a given set of inclusion probabilities π_i's, which is valid for the most general values of π_i's. Some of the advantages of this method are pointed out. In Section 3 we give a method of selecting units, one by one and independently, which has the advantage that we can stop at any stage we like, without distorting the proportionate values of π_i's at each stage. Some of these results were presented to the Indian Science Congress session at Bombay in 1960. Subsequently, Murthy (1960) gave another method of obtaining given set of π_i's by systematic sampling, and Hartley and Rao (1959) also dealt with the same problem. However, their methods are different.

2. Sampling scheme

For the present we shall ignore the necessity of having $\pi_{ij} > 0$ for all i and j and explain a method of obtaining the given values $\pi_1, \ldots, \pi_N$ for the inclusion probabilities. The slight modifications necessary to ensure the conditions $\pi_{ij} > 0$, $(i \neq j)$ will be given later.

Suppose without loss of generality, that

$$\pi_1 \geqslant \pi_2 \geqslant \pi_3 \geqslant \ldots, \geqslant \pi_N. \qquad \ldots \quad (2.1)$$

If this condition is not satisfied, we need only to rearrange members of (1.1) suitably. Let k_1 be the positive integer such that

$$S_1 = \sum_{i=1}^{k_1} \pi_i \leqslant 1 < S_1 + \pi_{k_1+1}. \qquad \ldots \quad (2.2)$$

Choose the units u_i in the first draw, with probabilities given by

$$p_i^{(1)} = \begin{cases} \pi_i & \text{if } 1 \leqslant i \leqslant k_1, \\ 1-S_1 & \text{if } i = k_1+1 \\ 0 & \text{if } i \geqslant k_1+2. \end{cases} \qquad \dots \quad (2.3)$$

We now choose $p_{k_1+1}^{(2)}$ such that the probability of including u_{k_1+1} in either of the first two draws equals π_{k_1+1} so that we can dispense with this unit in the subsequent draws. Clearly, we must have

$$p_{k_1+1}^{(2)} = \frac{\pi_{k_1+1}-(1-S_1)}{S_1}. \qquad \dots \quad (2.4)$$

As in (2.2), let k_2 be such that

$$S_2 = p_{k_1+1}^{(2)} + \sum_{k_1+2}^{k_1+k_2} \pi_i \leqslant 1 < S_2 + \pi_{k_1+k_2+1}. \qquad \dots \quad (2.5)$$

We then set

$$p_i^{(2)} = \begin{cases} 0 & \text{if } 1 \leqslant i \leqslant k_1 \\ \dfrac{\pi_{k_1+1}-(1-S_1)}{S_1} & \text{if } i = k_1+1 \\ \pi_i & \text{if } k_1+2 \leqslant i \leqslant k_1+k_2 \\ 1-S_2 & \text{if } i = k_1+k_2+1 \\ 0 & \text{if } i \geqslant k_1+k_2+2. \end{cases} \qquad \dots \quad (2.6)$$

We also see that we should have

$$p_i^{(n)} = 0 \quad \text{for} \quad 1 \leqslant i \leqslant k_1+k_2 \quad \text{and} \quad n \geqslant 3. \qquad \dots \quad (2.7)$$

The procedure can be continued till it terminates. If the last draw is r-th then we have

$$k = k_1+k_2+\dots+k_{r-1} < N$$

and there is a probability of

$$\epsilon = 1-(p_k^{(r)}+\pi_{k+1}+\dots+\pi_N) \qquad \dots \quad (2.8)$$

for not selecting any unit in the r-th draw. At every other draw, the probabilities assigned to the units add up to 1.

What is essentially done above is that we have ordered the units in a monotonic order of their sizes π_i, and divided the population into a number of homogeneous strata, which, however, are not strictly non-overlapping two adjacent strata having possibly a unit in common. (Such units will be called *junctional units* henceforth). We then selected one unit from each of these strata. The sampling in one stratum

is independent of that in any other stratum. The number of units in each stratum, and their sizes are so determined as to make the contributions to total of Y, of each stratum, nearly of the same order.

We notice that if u_i and u_j belong to the same stratum and neither of them is a junctional unit, then $\pi_{ij} = 0$, so that we cannot obtain an estimate of $V(\hat{Y})$ by our scheme of sampling. We can, however, modify as follows: we make the first draw with $p_i^{(1)} = \pi_i/\nu_0$, where

$$\nu_0 = \sum_{i=1}^{N} \pi_i.$$

Setting
$$\delta_i = \pi_i\left(1 - \frac{1}{\nu_0}\right) \div \left(1 - \frac{\pi_i}{\nu_0}\right) \qquad \ldots \quad (2.9)$$

it is easy to verify that if the probability of including u_i in the second or subsequent draws equals δ_i, then the total inclusion probability will be equal to π_i. Hence, all that is necessary is to proceed sampling for the second and subsequent draws with δ_i's as described above.

However, an alternative and perhaps a better method would be to draw two independent subsamples, both of the same expected size. If then π_i' denotes the inclusion probability for u_i in each subsample, it is easy to see that by setting

$$\pi_i' = 1 - \sqrt{1 - \pi_i} \qquad \ldots \quad (2.10)$$

the probability of including u_i in at least one of the subsamples becomes equal to π_i. Also, if π_{ij}' denotes the joint-inclusion probability of the pair (u_i, u_j) in a subsample, then

$$\pi_{ij} = 2(\pi_{ij}' + \pi_i'\pi_j') - \pi_{ij}'^2 \qquad \ldots \quad (2.11)$$

so that, knowing π_{ij}' we can easily calculate π_{ij}. The π_{ij}''s can be calculated using (1.5).

We shall examine, finally, the advantages of the above method. Even when the auxiliary values X's are exactly proportional to Y, it can be seen that $V(\hat{Y})$ where $\hat{Y}$ is the estimator as proposed by Horvitz and Thompson (1952) has still a component due to the variations in the effective sample size $\nu(s)$ Rao (1962). Hence a design should preferably have as small a variation in $\nu(s)$ as is possible. Further, the stability of the number of distinct units in the sample is an important guiding principle in practical operations, where cost should preferably be stable. The procedure given first, is then a good approximation since repetitions of units are minimised, only the junctional units having a small probability of getting repeated. These advantages are sacrificed to some extent in the two modifications given. It can be verified that there can be utmost 4 draws for each subsample, which give a positive probability to at least one of u_i and u_j, so that from (1.5), we see that the calculation of π_{ij} involves utmost 4 pairs of numbers. This fulfils an important practical requisite of quick and easy calculation of π_{ij}'s from $p_i^{(n)}$'s, especially if we deal with large samples. Further, by choosing a scheme of independent draws, we avoided the necessity of

tedious calculation of a large number of conditional probabilities of selection. Also, by utilising π_i's fully, we drastically reduced the labour of computing large number of probabilities. In fact, though we make 'r' draws (considering the first method given), out of the $r(N-1)$ independent probabilities to be used, we had only the $(r-1)$ junctional probabilities calculated, all the others being either zero or a π_i. When the population is large enough the gain in the labour of computations is quite considerable.

3. A SEQUENTIAL SCHEME

We shall now proceed to describe another scheme, which answers an important practical problem that arises frequently. In practice it is not the actual values of π_i's that are of interest to us, but only their proportionate values. Prior information on our auxiliary character, which is nearly proportional to the characteristic under consideration, being given, we would like to have a sampling scheme for which

$$\pi_i = KX_i, \qquad 1 \leqslant i \leqslant N,$$

so that

$$K = \frac{\Sigma\pi_i}{\Sigma X_i} = \frac{\nu_0}{\Sigma X_i} = \frac{\nu_0}{X}, \text{ say,}$$

where X_i is the value of X on u_i.

If we reasonably assume that the cost of inspecting a sample is proportional (roughly) to the number of distinct units in it, considerations of available resources fix an upper limit C_2 and considerations of desired accuracy fix a lower limit C_1, to the expected effective size of the sample, where, of course, $C_1 < C_2$. The value of ν_0 will then be chosen to be some convenient number between C_1 and C_2, and this fixes K. However, there will be situations when the cost per unit for the inspection of a sample cannot be estimated sufficiently accurately beforehand. Or there may be sudden changes resulting in a considerable difference between estimated costs and incurred costs. In such cases, it is desirable to have a sampling scheme which allows us a flexibility in the choice of number of units to be selected in the sample, even after the inspection of the sample, unit by unit, has started—and this it should be possible to do without distorting the basic property of the sampling design. In cases where we have information on an auxiliary character, we can choose as the basic property of the design, that it should have the inclusion probabilities π_i's proportional to the given X_i's. We now give a method of sampling which completely solves the problem in almost all important practical situations. This scheme consists of independent draws such that at the end of every draw (up to a certain stage), the inclusion probabilities attained till then, are proportional to a given set of values π_i, $(1 \leqslant i \leqslant N)$ say. Let $p_i^{(n)}$ be the probability of selecting u_i in n-th draw, and $\pi_i^{(n)}$ be the inclusion probability of u_i at the end of n-th draw. We want to have

$$\pi_i^{(n)} = K_n\pi_i \quad i = 1, 2, \ldots N; \quad n = 1, 2, \ldots \qquad \ldots (3.1)$$

where K_n's are constants. We can clearly assume that $\sum_{i=1}^{N} \pi_i = 1$, without loss of generality, and choose K_n's appropriately. At the end of n draws we have

$$\pi_i^{(n)} = 1-(1-p_i^{(1)})(1-p_i^{(2)})\ldots(1-p_i^{(n)}). \qquad \ldots (3.2)$$

Further, we have the conditions

$$\sum_{i=1}^{N} p_i^{(n)} = 1, \qquad n = 1, 2, \dots \qquad \dots \quad (3.3)$$

since the probabilities should add up to 1 at each draw. Setting $n = 1$, we have from (3.2) and (3.3),

$$K_1 = 1 \text{ and } p_i^{(1)} = \pi_i \quad (1 \leqslant i \leqslant N). \qquad \dots \quad (3.4)$$

Setting $n = 2$ in (3.2), we have for all 'i'

$$(1-p_i^{(1)})(1-p_i^{(2)}) = 1-K_2\pi_i$$

giving
$$p_i^{(2)} = \pi_i(K_2-K_1)/(1-K_1\pi_i). \qquad \dots \quad (3.5)$$

We then have from (3.3),

$$K_2 = K_1 + \left\{ \sum_{1}^{N} \frac{\pi_i}{1-K_1\pi_i} \right\}^{-1}. \qquad \dots \quad (3.6)$$

Similarly, setting $n = 3$ we have for $1 \leqslant i \leqslant N$,

$$(1-p_i^{(1)})(1-p_i^{(2)})(1-p_i^{(3)}) = 1-K_3\pi_i$$

giving
$$p_i^{(3)} = \frac{\pi_i(K_3-K_2)}{1-K_2\pi_i} \qquad \dots \quad (3.7)$$

and (3.3) then gives
$$K_3 = K_2 + \left\{ \sum \frac{\pi_i}{1-K_2\pi_i} \right\}^{-1}. \qquad \dots \quad (3.8)$$

In general (writing $K_0 = 0$ conventionally), it can be shown that

$$K_{n+1} = K_n + \left\{ \sum \frac{\pi_i}{1-K_n\pi_i} \right\}^{-1}. \qquad \dots \quad (3.9)$$

and
$$p_i^{(n)} = \frac{\pi_i(K_n-K_{n-1})}{1-K_{n-1}\pi_i} \qquad \dots \quad (3.10)$$

for $1 \leqslant i \leqslant N$ and for n upto a certain limit which we shall presently investigate.

From (3.1) and the assumption that $\Sigma \pi_i = 1$, we see that the expected effective sample size after the n-th draw, which is equal to $\Sigma_i \pi_i^{(n)}$, is equal to K_n. Hence K_n increases with n. When n tends to infinity we expect to get all the units of the population, so that

$$K_n \to N \quad \text{as} \quad n \to \infty.$$

Assuming as before that $\pi_1 \geqslant \pi_2 \geqslant \dots \geqslant \pi_N$, except in the degenerate case when all inequalities become equalities, we see from the assumption $\Sigma \pi_i = 1$, that

$$\pi_1 > \frac{1}{N}.$$

Hence there exists an integer n_0 such that

$$\pi_1 K_{n_0} \leqslant 1 \quad \text{and} \quad \pi_1 K_{n_0+1} > 1. \qquad \dots \text{(3.11)}$$

The procedure given above then yields in the (n_0+1)-th stage a negative probability for u_1, as can be seen from (3.10). Hence the procedure is valid upto the n-th draw only. However, we shall see that this is not a serious limitation in practice. If we wish to sample an expected effective proportion 'f' of the population, we wish to continue the procedure upto n-th draw, where

$$K_n \geqslant Nf > K_{n-1}. \qquad \dots \text{(3.12)}$$

Since we can continue the procedure up to n_0-th draw, as given by (3.11), we can have n draws if and only if

$$\pi_1 \leqslant \frac{1}{f} \cdot \frac{1}{N}. \qquad \dots \text{(3.13)}$$

In practice, f will be of the order of 5% to 10% or even less. Hence (3.13) is satisfied if the maximum of N positive quantities whose sum is equal to 1 is less than ten times the average. In practice this condition will in general be satisfied. However, if there are some units with exceptionally large π_i's we can perform independent binomial trials with the corresponding π_i's as the probabilities of successes, and after including the units for which successes are obtained, in the sample, we can eliminate all these units with large π_i's and proceed as before with the remaining units. A simpler and perhaps more practical method is to split each such large unit into 2 or more sub-units each having a small value of π_i and treat the selection of any single subunit as resulting in the selection of the corresponding entire unit.

We see then that at the end of every draw, the inclusion probabilities attained are proportional to given values of an auxiliary characteristic. If after drawing a sample of size n we want to increase the sample size, all we need to do is to perform the $(n+1)$-th, $(n+2)$-th,...draws with probabilities as given by (3.9) and (3.10), until we get the required size. Similarly, in cases where we are not quite sure of the reliability of our estimate of cost per unit of inspecting a sample, we should make some draws to get the sample and then inspect the units of the sample, one by one in the order in which they are selected—at least after a critical stage is reached. If at any stage we find that our resources are drying up or that sufficient accuracy is already expected to have been attained, we terminate the inspection at that stage and note the value of 'n' the number of draws made in all. The calculations involving π_{ij}'s are mathematically simple but computationally not as simple as they are in the procedures given in Section 2, but at times the labour is worth the advantages we are gaining. The procedure has attractive mathematical tractability (which is absent to some extent in the case of schemes given in Section 3) and allows us to draw samples of size upto 10% of the population and very often even more.

References

HANUMANTHA RAO, T. V. (1962) : An existence theorem in sampling. *Sankhyā*, Series A, **24**. 3, 327-330.

HARTLEY, H. O. and RAO, J. N. K. (1959) : Sampling with unequal probabilities and without replacement. *Proceedings of Social Statistics, Section of the American Statistical Association.* 56-59 (abstract).

HORVITZ, D. G. and THOMPSON, D. J. (1952) : A generalisation of sampling without replacement from a finite universe. *J. Amer. Stat. Ass.*, **47**, 663-685.

MURTHY, M. N. (1960) : A note on probability sampling. *Proceedings of the Fortyseventh Session of Indian Science Congress*, Part III ; 34 (abstract).

YATES, F. and GRUNDY, P. M. (1953) : Selection without replacement from within strata with probability proportional to size. *J. Roy. Stat. Soc.*, B, **15**, 253-261.

ON 'HORVITZ AND THOMPSON ESTIMATOR'

By T. V. HANURAV

Indian Statistical Institute

SUMMARY. In this paper we consider the problem of finding optimal sampling designs for the use of 'Horvitz and Thompson' estimator (1952) to estimate (unbiasedly, unless otherwise stated) the population total of a character Y, when auxiliary information on a correlated character X is available for all the units. Since there does not exist a design in which the variance is uniformly minimum, the optimal designs are obtained by minimising the expected variance under a realistic superpopulation set-up. These turn out to be designs in which the effective sample size is constant for all samples of the design. It is further proved that with the same criterion for comparison, the Horvitz-Thompson estimator for these optimal class of designs is uniformly superior to Des Raj's (1956) estimator in the symmetrised form for the sampling designs considered by him when the average effective sample size is 2.

INTRODUCTION

Consider a finite population, consisting of N units

$$u_1, u_2, \ldots, u_N. \qquad \ldots \ (1.1)$$

Let Y be a quantitative character, taking the value y_i (which is unknown, *a priori*) on u_i, $(1 \leqslant i \leqslant N)$. Let $D = D(S, P)$ be a *sampling design* consisting of a set S of samples s from (2.1), with a probability measure P defined on it. We define

$$\pi_i = \sum_{s \supset u_i} P_s, \qquad (1 \leqslant i \leqslant N). \qquad \ldots \ (1.2)$$

(where the summation extends over all samples s of S that contain u_i), to be the *inclusion probability* of u_i in D. Similarly, we define

$$\pi_{ij} = \sum_{s \supset u_i, u_j} P_s, \qquad (1 \leqslant i \neq j \leqslant N) \qquad \ldots \ (1.3)$$

to be the inclusion probability of the pair (u_i, u_j). An unbiased estimator of the population total

$$T = \sum_{i=1}^{N} y_i \qquad \ldots \ (1.4)$$

as proposed by Horvitz and Thompson, is then given by

$$\hat{T}_1 = \sum_{i \in s} \frac{y_i}{\pi_i} \qquad \ldots \ (1.5)$$

where the summation extends over all *distinct* units u_i belonging to s (i.e., we ignore repetitions). The variance of T_1 is given by

$$V(\hat{T}_1) = \sum_{i=1}^{N} y_i^2 \left(\frac{1 - \pi_i}{\pi_i} \right) + \sum_{i \neq j}^{N} \sum y_i y_j \left\{ \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j} \right\}. \qquad \ldots \ (1.6)$$

In many situations of practical interest, we have *a priori*, the values x_i, which another quantitative character X, highly correlated with Y, takes on u_i (for $1 \leqslant i \leqslant N$). In such cases, this information is used to construct sampling designs and estimators of, T say, which result in a greater precision than those which do not make use of this information. Examples of such procedures are pps estimator, Des Raj's

estimator, ratio estimator, regression estimator and so on. The amount of gain in precision due to these estimators depends on the degree of correlation between X and Y, and to assess this more fully, one needs to assume some broad statistical relationship between X and Y, so far as the population (1.1) is concerned.

When the value x_i is known for the unit u_i, it is reasonable to assume, in many practical cases, that the corresponding y_i (which, however, is unknown), is the outcome of a random variable Z whose expectation is proportional to x_i and whose variance is either partly or fully unknown. The realised value $y = (y_1, ..., y_N)$ can thus be considered as the realisation of N-length random vector from a super-population. This concept has been introduced by Cochran (1946) and since then has been successfully used by many others. We notice here that we tacitly make this assumption when dealing with pps estimator, ratio estimator, and almost all estimators used in designs of varying probability sampling, while when using regression estimator we make a similar but slightly weaker assumption. We explicitly formulate our model, writing E_1 and V_1 to denote the conditional expectation and variance, given x_i's, thus :

$$E_1(y_i/x_i) = a\, x_i \qquad \qquad \text{... (1.7)}$$

and
$$V_1(y_i/x_i) = \sigma_i^2 \quad \text{(say)} \qquad \qquad \text{... (1.8)}$$

where, a and σ_i^2's are unknown constants. We also assume that y_i and y_j are independent for all $i \neq j$, for given x_i and x_j. In particular this implies that

$$E_1(y_i y_j / x_i \text{ and } x_j) = a^2 x_i x_j. \qquad \qquad \text{... (1.9)}$$

2. OPTIMAL DESIGNS

Under the assumptions (1.7), (1.8) and (1.9), we shall proceed to find the sampling designs best suited for the use of $\hat{T}_1$ as given by (1.5), to estimate T. The criteria that we choose for the best are (1) unbiasedness and (2) minimum variance. Clearly, the increase of sample size, while increasing the precision of estimates, increases the cost also. Assuming that the cost of drawing and inspecting a sample s is proportional to the number of distinct units in the sample (which we shall call the 'effective size' of s and denote by $\nu(s)$, henceforth), we search for the best designs to use (1.5) in the class of all designs having a fixed given value for $\nu(s)$ for all s in the design. However, we may relax the later condition to include designs in which $\nu(s)$ can vary from sample to sample by demanding that the expected value of $\nu(s)$ be equal to a given value. This means that the expected cost of sampling is to be fixed, which is a reasonable condition unless the variation of $\nu(s)$ in our optimal design is too large. We note that

$$E\nu(s) = \sum_{s \in S} \nu(s) P_s = \sum_{i=1}^{N} \pi_i = \nu_0 \quad \text{say.} \qquad \qquad \text{... (2.1)}$$

Given the auxiliary information on X, we shall consider only those designs for which the inclusion probability π_i is proportional to $x_i (1 \leqslant i \leqslant N)$. (This is not only reasonable but is probably better as hinted in Section 3). Given the expected cost, ν_0 is fixed and our domain of search then becomes the class of all designs in which

$$\pi_i = c x_i \qquad (1 \leqslant i \leqslant N) \qquad \qquad \text{... (2.2)}$$

430

where c is a constant determined by (2.1) thus

$$c = \frac{v_0}{\sum\limits_{i=1}^{N} x_i} = \frac{v_0}{M}, \quad \text{say.} \qquad \ldots \text{(2.3)}$$

It is seen that in these designs, the variance of (1.5), as given by (1.6) depends on y_i's and π_{ij}'s. Hence it is the set of π_{ij}'s that are at our disposal in the choice of optimal designs. If there exists a subclass of the above designs, for which (1.6) is minimum, uniformly with respect to all possible values of $y = (y_1, \ldots, y_N)$, clearly, we have obtained the best designs. But we prove that (1.6) cannot be uniformly minimised with respect to y_i's. Setting

$$Q = \sum_{i=1}^{N} y_i^2 \left(\frac{1-\pi_i}{\pi_i} \right) + \sum_{i \neq j}^{N} \frac{y_i \, y_j}{\pi_i \, \pi_j} (\pi_{ij} - \pi_i \pi_j).$$

which is continuous in y's the conditions for Q to be minimum require

$$\frac{\delta Q}{\delta y_i} = 2y_i \frac{(1-\pi_i)}{\pi_i} + \sum_{\substack{j=1 \\ j \neq i}}^{N} y_j \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j} = 0$$

for all i and for all sets of y_i's. Clearly, this is violated by setting $y_i = 1$ for any fixed i and $y_j = 0$ for $j \neq i$. Hence we proceed to do the next best, which is to see whether the expected value of (1.6), when conditioned over given x_i's can be minimised, uniformly with respect to all values of x_i's, a and σ_i^2's. When even this is not possible, we then have to restrict our population of x_i's to some specific models. However, we shall prove that there exists an optimum class of designs with given π_i's for which the conditional expectation of (1.6) is uniformly minimised. From (1.7), (1.8), (1.9), (2.2) and (2.3), we have

$$E_1 V(\hat{T}_1) = \sum_{i=1}^{N} \left(\frac{1-\pi_i}{\pi_i} \right) (a^2 x_i^2 + \sigma_i^2) + \sum_{i \neq j}^{N} \left(\frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j} \right) a^2 x_i x_j$$

$$= \sum_i \left(\frac{1-\pi_i}{\pi_i} \right) \left(\frac{a^2 \pi_i^2}{c^2} + \sigma_i^2 \right) + \frac{a^2}{c^2} \sum_{i \neq j} (\pi_{ij} - \pi_i \pi_j). \qquad \ldots \text{(2.4)}$$

Hence $E_1 V(\hat{T}_1)$ is minimum, for given values of π_i's, a, c and σ_i^2's, when and only when

$$\sum_{i \neq j} \pi_{ij} \qquad \ldots \text{(2.5)}$$

is minimum. Defining auxiliary random variables R_{si} by

$$R_{si} = \begin{cases} 1 & \text{if } u_i \in s \\ 0 & \text{otherwise} \end{cases}$$

we have

$$\pi_i = \sum_{s \varepsilon S} R_{si} P_s$$

$$\pi_{ij} = \sum_{s \varepsilon S} R_{si} R_{sj} P_s$$

431

and
$$v(s) = \sum_{i=1}^{N} R_{si} = \sum_{i} R_{si}^2.$$

Hence
$$\sum\sum_{i\neq j} \pi_{ij} = \sum\sum_{i\neq j} \{ \sum_{s\epsilon} R_{si} R_{sj} P_s\}$$

$$= \sum_{s\epsilon S} \{ \sum_{i\neq j}^{N} \sum R_{si} R_{sj} P_s\}$$

$$= \sum_{s\epsilon S} \{v^2(s) - v(s)\} P_s$$

$$= v_0^2 + V(v(s)) - v_0. \qquad \ldots \ (2.6)$$

Hence, for a given v_0, the expected sample size, (2.6) is minimised when $V(v(s))$ is minimised. In other words, the design should contain, as far as possible, samples of the same effective size. When v_0 is an integer

$$\min \ V(v(s)) = 0$$

while if
$$v_0 = [v_0] + f, \qquad 0 < f < 1,$$

$[v_0]$ being the greatest integer not greater than v_0,

$$\min \ V(v(s)) = f(1-f) \qquad \ldots \ (2.7)$$

since we should have

$$v(s) = [v_0] \quad \text{with probability} \quad f$$

$$[v_0] + 1 \ \text{with probability} \ 1-f.$$

In practice, the considerable degree of freedom we have in the choice of v_0 allows us to have v_0 as an integer. Even otherwise, (2.7) is negligible in comparison with $(v_0^2 - v_0)$ for practical values of v_0.

The practical problem of the actual construction of these designs is not solved fully, but the author (Hanurav, 1962)* gave a method of selecting units from (1.1) which results in prescribed general values of π_i's. It is pointed therein that the resulting design has a very stable value of $v(s)$, serving as a good approximation for our purpose. We therefore assume that the minimum of (2.6), viz., $(v_0^2 - v_0)$ can be closely attained in practice, so that for purposes of comparison we can take

$$\min \ E_1 V(\hat{T}_1) = \frac{a^2}{c^2} \sum_{i=1}^{N} \pi_i(1-\pi_i) + \sum_{i=1}^{N} \left(\frac{1-\pi_i}{\pi_i} \right) \sigma_i^2 + \frac{a^2}{c^2} (v_0^2 - v^2 - \sum\sum_{i\neq j} \pi_i\pi_j)$$

$$= \frac{a^2}{c^2} \left(v_2 - \sum \pi_i^2 \right) + \sum \left(\frac{1-\pi_i}{\pi_i} \right) \sigma_i^2 + \frac{a^2}{c^2} \{v_0^2 - v_0 - (v_0^2 - \sum \pi_i^2)\}$$

$$= \sum \left(\frac{1-\pi_i}{\pi_i} \right) \sigma_i^2 \qquad \ldots \ (2.8)$$

*Published in the same issue, pp. 421-428.

3. COMPARISON WITH SYMMETRISED DES RAJ'S ESTIMATOR

For the purpose of this section, we shall make one further assumption beside (1.7) and (1.9), and this is regarding the conditional variances σ_i^2's. A commonly advocated assumption is that σ_i^2's are all equal but unknown. However, in many cases of practical interest (especially when the variates Y and X are positive as is the case in most of the sample surveys) it is more realistic to assume that the conditional coefficients of variation are more or less same for all units so that the conditional variance increases with the conditional mean. We explicitly write this as

$$V_1(y_i/x_i) = K \cdot a^2 x_i^2 = \sigma^2 x_i^2 \qquad \dots \ (3.0)$$

where σ^2 is an unknown constant.

We now compare the estimator $\hat{T}_1$ used in the optimal class of designs derived above, with the symmetrised Desraj's (1956) estimator, when $\nu_0 = 2$, under the assumptions (1.7), (1.9) and (3.0). We briefly describe this later estimator.

We draw a sample size n say, without replacement. At each draw, the probability p_i of selecting u_i is proportional to x_i, if it is not already selected. Here p_i will be taken proportional to x_i, where x_i has the same meaning as in Sections 1 and 2. If u_i is selected in the first draw, the probability of selecting u_j in the second draw is

$$p_j^{(2)}(i) = \begin{cases} \dfrac{p_j}{1-p_i} & \text{if } j \neq i \\ 0 & \text{if } j = i. \end{cases}$$

Similarly, in the third draw we have

$$p_k^{(3)}(i,j) = \begin{cases} \dfrac{p_k}{1-p_i-p_j} & \text{if } k \neq i, k \neq j \\ 0 & \text{otherwise} \end{cases}$$

where the notations are clear. An unbiased estimator of T in such cases, as given by Desraj (1956) is

$$\hat{T}_{2,\text{asym}} = \frac{1}{n} \sum_{a=1}^{n} t_a \qquad \dots \ (3.1)$$

where, $$t_a = y_{i_1} + y_{i_2} + \dots + y_{i_{a-1}} + \frac{y_{i_n}}{p_{i_n}} (1 - p_{i_1} - p_{i_2} \dots - p_{i_{a-1}}),$$

$u_{i_1}, u_{i_2}, \dots, u_{i_{n-1}}, u_{i_n}$ being the units successively obtained in the sample, the suffix "asym" denoting that the estimator is asymmetric in the observed values. Restricting ourselves to an important practical case of $n = 2$, we write (3.1) thus

$$\hat{T}_{2,\text{asym}} = y_i + \frac{y_j}{p_j} (1 - p_i) \qquad \dots \ (3.2)$$

where u_i is selected in the first draw and u_j in the second draw. It is well known that this estimator can always be improved by taking the weighted mean of different asymmetric estimators, for given unordered sample, the weights being the respective conditional probabilities of obtaining the ordered samples given the unordered sample (Halmos, 1946). Denoting this improved symmetric estimator by $\hat{T}_2$, we have

$$\hat{T}_2 = \frac{1}{2-p_i-p_j} \left\{ (1-p_j) + \frac{y_i}{p_i}(1-p_i)\frac{y_j}{p_j} \right\} \qquad \ldots \ (3.3)$$

which is symmetric in i and j as it should be. We have

$$V(\hat{T}_2) = \mathop{\Sigma\Sigma}_{\substack{i \neq j \\ 1}}^{N} p_i p_j \left(\frac{1-p_i-p_j}{2-p_i-p_j} \right) \left[\frac{y_i}{p_i} - \frac{y_j}{p_j} \right]^2$$

$$= \mathop{\Sigma}_{i=1}^{N} \frac{y_i^2}{p_i} \left\{ \mathop{\Sigma}_{j \neq i} p_j \frac{(1-p_i-p_j)}{(2-p_i-p_j)} \right\} - \mathop{\Sigma\Sigma}_{\substack{i \neq j \\ 1}}^{N} y_i y_j \left\{ \frac{1-p_i-p_j}{2-p_i-p_j} \right\}. \qquad \ldots \ (3.4)$$

In order to compare $\hat{T}_1$ and $\hat{T}_2$ we should take π_i's for $\hat{T}_1$ such that $\nu_0 = 2$, so that the expected effective sample size remains the same in both cases. Since $\mathop{\Sigma}_{1}^{N} p_i = 1$, and both p_i's and π_i's are proportional to x_i's,

$$p_i = \frac{\pi_i}{2}, \qquad \ldots \ (3.5)$$

so that from (1.7), (1.9), (2.8), (3.0), (3.4) and (3.5),

$$E_1 V(\hat{T}_2) = \mathop{\Sigma}_{i=1}^{N} \frac{2}{\pi_i} \cdot \frac{\pi_i^2(a^2+\sigma^2)}{c^2} \left\{ \mathop{\Sigma}_{j \neq i} \frac{\pi_j(2-\pi_i-\pi_j)}{2(4-\pi_i-\pi_j)} \right\} - \mathop{\Sigma\Sigma}_{i \neq j} \frac{a^2}{c^2} \pi_i \pi_j \left[\frac{2-\pi_i-\pi_j}{4-\pi_i-\pi_j} \right]$$

$$= \frac{\sigma^2}{c^2} \mathop{\Sigma\Sigma}_{i \neq j} \pi_i \pi_j \left[\frac{2-\pi_i-\pi_j}{4-\pi_i-\pi_j} \right]. \qquad \ldots \ (3.6)$$

We assume that the minimum variance of $\hat{T}_1$ as given by (2.8) can be closely attained and that we can neglect the component of variance arising due to the slight variations in the effective sample size. Hence we shall proceed to compare (3.6) and (2.8). We have

$$E_1 V(\hat{T}_2) - \min E_1 V(\hat{T}_1)$$

$$= \frac{\sigma^2}{c^2} \left[\mathop{\Sigma\Sigma}_{i \neq j} \frac{\pi_i \pi_j(2-\pi_i-\pi_j)}{(4-\pi_i-\pi_j)} - 2 + \Sigma \pi_i^2 \right]$$

$$= \frac{2\sigma^2}{c^2} \left[1 - \mathop{\Sigma\Sigma}_{i \neq j} \frac{\pi_i \pi_j}{(4-\pi_i-\pi_j)} \right]$$

$$= \frac{2\sigma^2}{c^2} \left[1 - 2 \mathop{\Sigma\Sigma}_{i \neq j} \frac{p_i p_j}{(2-p_i-p_j)} \right]. \qquad \ldots \ (3.7)$$

Let
$$\chi(p_1, \ldots, p_N) = \sum_i \sum_j \frac{p_i p_j}{(2 - p_i - p_j)}.$$

When
$$p_i = \frac{1}{N} \text{ for all } i,$$

clearly
$$\chi\left(\frac{1}{N}, \frac{1}{N}, \ldots, \frac{1}{N}\right) = \frac{1}{2},$$

so that in this case
$$E_1 V(\hat{T}_2) = \min E_1 V(\hat{T}_1).$$

In order to prove that
$$E_1 V(\hat{T}_2) \geqslant \min E_1 V(\hat{T}_1),$$

we shall prove that χ is actually maximum when all its arguments are equal to $1/N$. We have the restriction
$$\Sigma p_i = 1$$

on the p_i's. Introducing the Lagrangian multiplier λ,

let
$$\psi = \sum_i \sum_j \frac{p_i p_i}{(2 - p_i - p_j)} - \lambda(\Sigma p_i - 1).$$

We can verify that at the point where $p_i = \frac{1}{N}$ for $1 \leqslant i \leqslant N$, we have

$$\frac{\delta \psi}{\delta p_i} = 0$$

$$\frac{\delta^2 \psi}{\delta p} = \frac{-N(2N-1)}{4(N-1)^2} = -b_1 \quad \text{say}$$

$$\frac{\delta^2 \psi}{\delta p_i \, \delta p_j} = \frac{N \, N^2 + (N-1)^2}{4(N-1)^3} = b_2 \quad \text{say}$$

so that the Hessian of ψ is given by the $N \times N$ determinant

$$H(\psi) = \begin{vmatrix} -b_1 & b_2 & b_2 & \cdots & b_2 \\ b_2 & -b_1 & b_2 & \cdots & b_2 \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ b_2 & b_2 & b_2 & \cdots & -b_1 \end{vmatrix}$$

The value of the r-th order principal minor of $H(\psi)$ is
$$(-1)^{r-1}(b_1 + b_2)^{r-1}[(r-1)b_2 - b_1]. \qquad \cdots \quad (3.8)$$

435

14

Since $b_2 < b_1$, it follows that (3.8) changes its sign alternately as r increases and that it is negative when $r = 1$. This shows that χ attains its maximum when all p_i's are equal to $1/N$ and hence it follows that

$$\min E_1 V(\hat{T}_1) \leqslant E_1 V(\hat{T}_2)$$

which shows that when we average over the conditional variations of y_i's, $\hat{T}_1$ is uniformly superior to the symmetrised Desraj's estimator, when samples are of average effective size 2.

Remark : The above result justifies the opinion that when auxiliary information X of the type discussed above is available, it is preferable to choose the sampling scheme so as to make the inclusion probabilities π_i's proportional to x_i's instead of choosing the design with probability of selection in each draw proportional to x_i's. We note the assumption involved in taking (2.8) to be the minimum attainable variance when we use (1.5). This amounts to assuming that the given set of π_i's can be partitioned into two subsets such that in each subset, the total of the π_i's is exactly equal to unity. This assumption though need not hold good in general is a good approximation in practical cases especially when π_i's are small quantities as is usually the case. It also seems reasonable to conjecture the validity of our result even when $\nu_0 > 2$.

REFERENCES

COCHRAN, W. G. (1946) : Relative accuracies of systematic and stratified random samples for a certain class of population. *Ann. Math. Stat.*, **17**, 164-179.

DES RAJ (1956) : Some estimators in sampling with varying probability without replacement. *J. Amer. Stat. Ass.*, **51**, 269-284.

HALMOS, P. R. (1946) : The theory of unbiased estimation. *Ann. Math. Stat.*, **17**, 34-43.

HANURAV, T. V. (1962) : Some sampling schemes in probability sampling. *Sankhyā*, Series A, **24**, 421-428.

HORVITZ, D. G. and THOMPSON, D. J. (1952) : A generalization of sampling without replacement from a finite universe. *J. Amer. Stat. Ass.*, **47**, 663-685.